# POSSIBILISTIC PROCESSES FOR COMPLEX SYSTEMS MODELING

BY

**Cliff Alan Joslyn**

B.A., Oberlin College, 1985
M.S.A.T., State University of New York at Binghamton, 1989

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Advanced Technology
with Specialization in Systems Science
in the Graduate School of the
State University of New York
at Binghamton
1994

Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Systems Science
in the Graduate School of the
State University of New York
at Binghamton

George J. Klir, Chair ______________________ April 11, 1994
Department of Systems Science

Howard H. Pattee ______________________ April 11, 1994
Department of Systems Science

Valentin F. Turchin ______________________ April 11, 1994
Department of Computer Science, City College of New York

## 0.1 Abstract

**Possibility theory** is being developed as an alternative to traditional information theory. While possibility theory is logically independent of **probability theory**, they are related: both arise in **Dempster-Shafer evidence theory** as **fuzzy measures** defined on **random sets**; and their distributions are both **fuzzy sets**. Together these fields comprise the new field of **Generalized Information Theory** (GIT).

Traditionally mathematical possibilistic semantics has been based *strictly* on fuzzy sets and their interpretation in the context of psychological uncertainty and subjective evaluations. The purpose of this dissertation is to extend interpretations and applications of possibility theory beyond those of fuzzy sets; in particular, to develop a **natural semantics** of possibility for the purposes of **qualitative modeling** of **complex physical systems**. The dissertation addresses the following:

**Possibility Theory in GIT:** The relations between possibility theory and the other formalisms of GIT are explicated; **random set distributions** and their **distribution operators** and **structural** and **numerical aggregation functions** are introduced to relate probability with possibility in the context of GIT; possibility arises from **consistent random sets**; and methods for **possibilistic normalization** and **possibilistic approximation** of inconsistent random sets are developed. It is argued that there is no special relationship between possibility theory and fuzzy systems theory.

**Semantics of Possibility:** Drawing from **semiotics** and general **models**, criteria for the **natural semantics** of possibility are explored; the basis for a **graduated, *de re*** possibility is related to **modal**, **natural language**, and **probabilistic** views; a **strong compatibility requirement** for possibility and probability is advanced; **possibilistic concepts** are developed from **mathematical**, **statistical** and **physical** interpretations; and the **traditional semantics** of possibility from **subjective evaluations**, **converted probabilities**, and **likelihoods** are critiqued.

**Possibilistic Measurement:** Measurement methods for possibility values based on **subset observations**, and which are consistent with possibilistic semantics, are developed; **possibilistic histograms** which are **fuzzy intervals**, and their **continuous approximations**, are defined; set statistics are derived from **indirect measurement** of system components, **ensembles** of differently calibrated instruments, **interval-based time series data** from **order statistics**, and **local extrema** of time series data.

**Possibilistic Processes:** **General processes** are defined as **semirings** operating on **state vectors** and **transition matrices**, and the special case of **possibilistic processes** using max/t-norm semirings and possibilistically normal **conditional transition matrices**, are introduced, and their properties developed; **possibilistic Markov processes** and a **possibilistic Monte Carlo method** are defined.

**Software Architecture:** An **architecture** for a C++ implementation of possibilistic and GIT methods is proposed in the context of the **Computer-Aided Systems Theory** (CAST) research program.

**Qualitative Model-Based Diagnosis and Trend Analysis:** The use of possibility theory as a new method for **qualitative modeling** is explored. The potential for the application of possibilistic methods in systems for the qualitative **model-based diagnosis** and **trend analysis** of complex systems like spacecraft is described.

## 0.2 Acknowledgments

It is rare for someone to create a new field of knowledge. This work is certainly not an example of such a thing. Instead, it merely reflects the continuation in a certain direction of the work of someone who has, my teacher, Prof. George Klir of SUNY-Binghamton Systems Science. Without his vision and strong support over many years none of this would have been possible. Studying under him has been an immense privilege.

In my many years working at SUNY, the other pole of my intellectual life has been Prof. Howard Pattee. Before entering SUNY, I could not have imagined how rich and broad my study under Prof. Pattee would be, and how far my education would advance at the hands of such an inspiring teacher.

In addition, I would like to acknowledge the assistance of the other Systems Science faculty, and also Ms. Bonnie Cornick, without whom I would have been even more lost and confused than I normally was.

Among all the teachers I have had the pleasure of studying under, there is none I wish to thank more than Prof. Valentin Turchin of City College Computer Science. Since meeting him many years ago, all my work with him has been an unfolding process of discovery and wonder. It is rare indeed to find a teacher and friend of such brilliance and depth.

Our work together with Dr. Francis Heylighen of the Free University of Brussels is a continuing source of great satisfaction. I am all the more fortunate to have *two* colleagues with whom I share such a firm view of the world and close collaborative relationship. I know we all look forward to the work for the years ahead.

The last three years of my graduate work were funded by the Code 520 group of NASA's Goddard Space Flight Center in Greenbelt Maryland under grant # NGT 50756. I'm pleased to extend my deep gratitude to Dr. Walter Truszkowski for his support and collaboration. My work with the Goddard group is reflected not only in the final chapter, but throughout the dissertation. It would indeed have been impossible to complete this work without him, and I hope that he is pleased with the chance he took in supporting me. In addition, the whole Code 520 group, especially Michael Moore, were very helpful, as was the entire University Affairs office and staff at Goddard.

My good fortune actually began before entering SUNY, since I came with many of the intellectual resources that would serve me over later years. For this I must credit the superb education I received at Oberlin College, and especially the incredible support of my primary teachers there, Prof. Dan Merrill and Prof. Christian Koch. It was through my work with them that I was first able to explore the life of the mind, and to begin to realize the direction of my work. I will never forget the respect, freedom, and guidance they gave a young student.

Possibility theory is a strange and new enough field that there are few proper colleagues. My recent cooperation with Prof. Etienne Kerre of the University of Ghent and his students Dr. Gert de Cooman, Mr. Bernard de Baets, and Dr. Bart Cappelle has been especially welcome. I was delighted to find some *real* mathematicians trying to move this theory along! In addition Profs. Didier Dubois and Henri Prade of the Paul Sabatier University in Toulouse have been very kind and helpful.

The colleagues and friends I have made in the academic community over the years have also been a great source of support and stimulation. There are too many to mention them all. But I won't let that stop me (in alphabetical order, no less!): thanks to Prof. Pierre Auger, Prof. Bela Banathy, Dr. Kirstie Bellman, Dr. Peter Cariani, Prof. Francois Cellier, Dr. John Dockery, Prof. Paul Fishwick, Dr. Sally Goerner, Dr. Irwin Goodman, Dr. Kevin Hufford, Dr. George Kampis, Dr. Kevin Kreitman, Dr. Robert Lea, Prof. Hal Linstone, Prof. Alvaro Moreno, Prof. Tuncer Ören, Prof. Franz Pichler, Dr. William Powers, Mr. Luis Rocha, Prof. Stan Salthe, Prof. Len Troncale, Dr. (yet?) Jon Umerez, Prof. Stuart Umpleby, and (last but not least) Dr. Jack Wang for their friendship and help.

Finally, I would like to thank my family for believing in me through the years: Mom and Dad, Lisa and the folks in Maine, Sara, Paul, and Jonathan, and Otto and the big Mi.

*Cliff Joslyn*
*Spring 1994*
*Portland, Maine*

## 0.3 Preface

In 1983 I was a Sophomore at Oberlin College facing the prospect of declaring a major and planning the rest of my college career. After considering the possibility of completing a less than superlative mathematics program, I decided to take advantage of Oberlin's strong independent study and individual major programs. With the help of Profs. Dan Merrill and Chris Koch, and some long hours in the library, I designed an individual major program in Cognitive Science, the first formal study ever done in that subject at Oberlin.

It was during this time that I encountered the work of Gregory Bateson, Valentin Turchin, Kenneth Sayre, and Ludwig von Bertalanffy, and immediately realized where my true interests lay. In my senior year I deflected my studies towards Systems Science and Cybernetics, and have never looked back. Again with the unparalleled support of my teachers, I completed an honors thesis entitled "Cybernetics and the Science of Mind", and was graduated with high honors.

In that work I tried to show the validity and feasibility of Systems Science and Cybernetics as a research program for Cognitive Science. In doing so, I relied heavily on probabilistic information theory and entropic concepts to serve as a universal modeling language, available at all levels of analysis, and allowing the linkage of theories across qualitatively distinct types of systems. Similarly, semiotics and the general application of concepts from semantics provided a bridge from information theory to the special sciences of biology and psychology.

A few years after leaving Oberlin I entered graduate study in the SUNY-Binghamton Systems Science program. Now with the strong guidance of my new teachers Prof. George Klir and Prof. Howard Pattee, I had the perfect opportunity to absorb the contents of Systems Science and Cybernetics, in both its breadth and depth. And as I entered the doctoral program, I again faced the prospect of declaring a research topic and planning the rest of my graduate career. After some meditation, there was no question about the direction I would take.

Systems Science is highly variegated and heterogeneous, both synthetic and syncretic. More than any other field of study, it faces the fundamental dilemma of providing for unity amid diversity. From electrical engineering to family therapy to theoretical biology to constructivist philosophy, there are many roads to Systems Science. And my road has always been the search for a universal modeling language through information theory and semiotics.

From Prof. Klir I had been introduced to the expanding horizons of Generalized Information Theory (GIT), in particular the elegant symmetry between probabilistic and possibilistic information theory. And while investigating possibility theory, I realized that outside of the formalisms themselves, there had been very little effort to relate possibility theory (or fuzzy theory, evidence theory, or any of the other branches of GIT) to the kinds of foundational issues of systems science that had occupied me earlier: the availability of a universal modeling language to link type-specific descriptions of systems. These other formalisms had been developed *solely* in the context of modeling human psychology. And not only were their advocates *content* with that situation, they strongly *championed* it.

My path was clear: to explore the semantics of possibility with respect to the modeling of physical systems; in particular, to develop possibilistic models with valid methods for measurement and prediction of possibilistically distributed values.

Along the way a number of other goals became associated with this work. I found that possibilistic models and the empirical semantics of possibility theory were so poorly developed that I would have to start essentially at the beginning, defining the most basic measurement procedures for possibility distributions and the concepts of possibilistic processes from first principles.

I also found that random set theory provided the strongest base, both formally and at the semantic level, from which I could approach possibility theory. This view led to a strengthening of my criticism of the traditional mathematical and semantic understanding of possibility theory, which is based on fuzzy sets. It also indicated the direction to be taken to develop empirical measurement procedures for possibility.

I have been accused by some of committing the sin of the mathematician: of working from formalism to application, rather than from a problem to its solution. I have never understood this orthodoxy, and must admit that to a certain extent I have been "shopping for a problem". But fortunately I think I've found one in the developing methodology of qualitative modeling and simulation. I am working with Walter Truszkowski at NASA–Goddard to apply possibilistic methods to the qualitative model-based fault diagnosis and trend analysis of spacecraft systems, and we are both hopeful about the future of this work.

Finally, the search for a universal modeling language extends beyond formal mathematics to the implementations of these mathematical systems in computer-based languages. The Computer-Aided Systems Theory (CAST) school reflects the aspirations in this direction. It is my strong hope that I will be able to participate in the development of CAST systems to implement possibilistic and other GIT methods. If these attempts are well-engineered, then the resulting systems should provide a valuable basis for the development of higher-level Systems Science methodologies, for example Klir's General Systems Problem Solver (GSPS).

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Science and Information

I have come to believe that the whole world is
an enigma, a enigma that is made terrible by
our own mad attempt to interpret it as though
it had an underlying truth.

— *Umberto Eco*

This work is ultimately about a particular manifestation in the modern world of an ageless dualism present in the Western tradition: the distinction between the objective and the subjective; between the material and the immaterial; between phenomena and the referential semantic systems with which we model and interpret them; between physics and mathematics; between the physical and the mental; indeed, between body and mind.

The manifestation of these dualisms which will concern us here is the changing nature of the relation between the (broadly conceived) **natural sciences** and the **informational sciences**. Recently there have been great advances in the informational sciences, and a plethora of new formalisms. Whereas traditionally information theory has been tightly coupled to the natural sciences, whatever connection there might be between these new informational formalisms and the science of natural or physical systems has not generally been explored. In particular, it is the purpose of this thesis to develop the mathematical and methodological ideas which will help establish the new theory of **possibilistic information**, and to apply it to the modeling of physical systems.

## 1.1 Informational Properties and Concepts

The term "information science" has come to be used primarily to refer to what used to be called "library science". Today it encompasses the general problems of technological information management and retrieval systems. However, it will be used here to refer to the whole array of formal and semi-formal theories which address not the physical, but specifically the *informational* properties of systems. Thus information science as used here includes not only classical **information theory** and the new **generalized information theory** (GIT) which is of specific interest in this thesis, but also the broader fields of **computer science, linguistics** and **semiotics**, and parts of **philosophy of science** and **philosophy of language**.

In order to try to achieve some coherence in what can be a semantic minefield, we begin by explicating some of the basic concepts which surround the term "information".

### 1.1.1 Fundamental Characterizations

The essential requirement for the nominal concept of information is actually a *process* of "inform"-ation — a process by which something becomes **informed**. This creates an image of the reception of a message, a "piece" of "knowledge" by which we become informed about what the actual state of affairs is, and thus an increase in the "amount" of that knowledge. This can also be seen as the resolution of a question, as the selection of an answer from an ensemble of possible answers.

It is usually held that an increase of information, a resolution of a question, results in a corresponding *decrease* of something else: of the capacity to store *future* information, of the number of questions *yet to be* answered, or of the number of *possible* states of affairs.

As noticed by Smithson [271], traditionally it has only been the first of these concepts which is held in a positive manner, while the other is usually seen *only* as a *lack* of information or certainty. There are a host of terms with a negative prefix to describe this lack of information, such as **uncertainty, imprecision, incompleteness, ignorance, nonspecificity, indefiniteness, indistinctness, indeterminacy, inexactness**, and **unclarity**. There are correspondingly few unmodified positive terms, such as **doubt, randomness, ambiguity** and **vagueness**.[1]

But ultimately the presence of information and uncertainty are inherently interrelated, and must always be mutually present. The reception of information necessarily involves the elements of **surprise** and **novelty**, a change that could not be foreseen, that was unknown before the reception of the information. The occurrence of novelty or surprise necessarily requires an inherent **freedom** in the system in which the novelty occurs, which freedom is then reduced by the reception of the

[1] It is interesting to note that in classical information theory the situation is actually reversed, where the concept of **entropy** represents a lack of knowledge, which prompted Schrödinger [257] to coin the term **negentropy** to describe the presence of form.

information.

In the end, following Ashby [6], the two dual, fundamental concepts of **constraint** and **variety** are arrived at as the foundations of a theory of information. In all of the above formulations, an increase in information is an increase in constraint, of a "giving of form", and a corresponding decrease of variety. Of course, information could also be lost, which would result in an increase in variety. These are both very general concepts, but have the advantage of being able to be treated mathematically in a clear way, having little connotative baggage, and both being positive terms.

Thus the task of formal information theory becomes that of trying to explicate the various ways that variety and constraint can be characterized, and whatever laws govern their change in systems.

### 1.1.2 Related Concepts and the Rejection of "Information"

There are unfortunate problems with the use of the term "information" itself in formal information theory, and particularly with the sense we are adopting here. Throughout its history, controversies have raged about the appropriateness of the term to describe a measure of constraint, as discussed by Shannon [263]. Information has been taken to mean the opposite of this sense, for example in the classical communication theory of Shannon and Weaver [264], where it is actually associated with a quantity of *variety*, in terms of the capacity of a channel to communicate information.

Huge confusions also exist in the literature about the distinction between the syntactic concept of information as a measure of variety or constraint and the corresponding semantic concepts of information which relate to **meaning, value, significance**, and other referential concepts in semiotics and linguistics. While the former is very exact, the latter can be quite problematic. Many wise people have chosen to avoid this issue by avoiding the term "information" itself, resorting to uncertainty or some more specific term. This will be aspired to here.

The concept of **indeterminism** and its relation to uncertainty is also important. Indeterminism is typically held to be an ontological property of some process in a phenomenal system, whereas uncertainty is an epistemic property of some modeling system. Thus the presence of indeterminism in a system would entail uncertainty in a good model of that system, and the uncertainty in that model would reflect the indeterminism in the object system. But on the other hand, the presence of uncertainty in a model does not entail a corresponding indeterminism in the object system: we may simply be ignorant of the determinism nevertheless present in the object process. Since our knowledge is necessarily restricted to the models which we can construct, the presence of indeterminism as an ontological property of real systems thus becomes a metaphysical question.

The recent growth in information science has also brought together a constellation of other concepts which require careful consideration. These concepts, including **order, organization, randomness**, and **complexity**, have been used with somewhat reckless abandon in recent years. In general they will not be used in this work except in passing, although it is suggestive to consider how the new information theories would relate to these and other concepts.

## 1.2 Natural Science and Information Theory

For centuries classical information theory developed in intimate relation to existing theories of physical systems, and has come to dominate many modern physical theories.

And with the advent of **programmable machines** in the mid-20th century, information science gained a new branch in the theory of **computational systems**. This involved not only theories about the nature of systems which perform computations and simulations which they carry out, but also theories about the fundamental limits of computation, and resulting limits on our knowledge of the natural systems they model. And increasingly, these computational information theories are also finding strong relationships to both classical and quantum theories of physical systems.

### 1.2.1 Probability

The history of information theory can be traced back to the first attempts to formalize theories of uncertainty in terms of the theories of chance and probability by Leibniz, Bernoulli and Laplace. In fact, probability is at the core of the centuries of development of classical information theory. It also has a key role to play in GIT, and has a strong relationship to possibility theory. The similarities and differences, both formal, philosophical, and semantic, between probability and possibility will occupy much of this work.

One of the foundational philosophical issues in probability is the distinction between aleatroy, *de re*, objectively-determined and *de dicto*, subjectively-determined probabilities [112]. This issue is also relavent for the modern information theories which are not specifically stochastic. Since our interest in determining of *de re* or aleatory *possibility* values will depend on some relation between a possibilistic measurement procedure and a *physical* system being measured, this issue will also be a fundamental concern for possibility theory.

Other issues in the foundations of probability concern its additivity and the representation of ignorance through the Principle of Insufficient Reason. These issues were decided in a specific way in previous centuries. But Shafer [262], Hacking [112], and others have discussed that there were many other ways these issues others might have been decided. The establishment of GIT today in part is a response towards the realization of these other possibilities.

### 1.2.2 Thermodynamic Information

Classical mechanics left no room for either uncertainty or indeterminism in systems having dynamical descriptions. This world had no capacity for novelty or surprise: from the universal Hamiltonian all would be revealed. Of course all was not quite well with this view. Almost all interesting differential systems, including those which model three astronomical bodies or a mole of gas, yield no analytical solutions, only sloppy approximations; and frictional forces were a constant nagging problem.

Explicit attempts to alleviate these problems began with the development of statistical physics by Boltzmann and Gibbs in the late 19th century. Statistical physics placed probability at the center of the physics of macroscopic phenomena. The long-term evolution of macroscopic, measurable quantities was modeled by the asymptotic behavior of probability distributions. In this way, certainty about the long-term behavior of a small number of gross aggregates of matter was traded for uncertainty about the short-term development of vast numbers of simple entities, and friction could be accommodated. Weaver [300] describes this in terms of informational concepts as the movement to a science of "disorganized complexity".

It was in the context of thermodynamics and statistical physics that the concept of **entropy** arose, originally in thermodynamics as a measure of the "thermodynamic distance" of a system from equilibrium, and later in statistical physics as a mathematical measure of the variety present in the probability distribution of the various energy levels of the mechanical states of the system. Where the former is a physical, ontological concept wedded to the context of the physical system being measured, the latter is an informational, epistemic concept tied only to the mathematics of probability distribution in question. So it is here that the rise of informational concepts in direct relation to the theories of physical systems can first be clearly seen

This relationship has continued throughout the further developments of thermodynamics. Whereas thermodynamic theories had been developed for closed or isolated systems, the thermodynamics of open systems came to the fore in later decades, beginning with the general theories of von Bertalanffy [12] and the near-equilibrium physics of Ilya Prigogine [221, 222]. Schrödinger [257] observed that the thermodynamics of open systems was intimately related to the fundamental processes of living systems. Later workers have furthered this work into speculations about fundamental relations between thermodynamic processes and processes of general evolution and emergence [20, 140, 292, WeBDeF88]. The key idea to these theories is that general evolution is in part a process by which systems export entropy to their environments, thus maintaining themselves in low entropy, highly informed, highly structured, "far-from-equilibrium" states.

### 1.2.3 Information Theory

In the rise of thermodynamics informational concepts were advanced into *universal* characteristics of systems at all levels of analysis, but still only in the *one* context of their physical, energetic thermodynamics. Thus concepts of thermodynamic entropy become significant to, for example, social and economic systems, to the extent that they involve physical processes of energy flow.

But beginning with the communications theory of Shannon and Weaver [264] in the 1950's and 1960's, a divergent view began to take hold which divorced entropy from its purely thermodynamic grounding into a general measure of variety in systems irrespective of their energetics. This view placed the emphasis on entropy as a measure of variety in stochastic systems, and therefore applied to stochastic systems described at any level and in any terms.

The ensuing debate in the literature about the relation between thermodynamic and nonthermodynamic entropies has yet to abate. Our purpose here is not to pursue this issue in depth.[2] No doubt there is a significant relation between entropy as a measure of a communications versus a thermodynamic system, as first argued by Brillouin [19] and Szilard [279]. But from the 1960's on there has been a proliferation of attempts to apply non-thermodynamic entropies in virtually every scientific context imaginable, from genetics to anthropology. Many researchers were attracted by what can turn out to be purely metaphorical relationships to such concepts as order and organization discussed above. Others of these efforts were not very well grounded conceptually or formally, as Shannon himself was the first to acknowledge [263].

But in the successes of this effort the continued growth of the significance of

[2] The reader is referred to a more thorough analysis of the issues involved in a paper by the author [128].

the informational sciences for the natural sciences can be seen. As distinct from the general thermodynamic view, with this approach informational concepts can be applied in a type-specific manner to systems at any level of analysis, opening the possibility of fruitful information analyses of systems in their specificities. A recent cogent example is the work of Tom Schneider [255,256], in which Shannon's coding theory is applied to the physics of "macromoleculer machines" such as enzymatic reactions and DNA transcription.

### 1.2.4 Quantum Information

Whereas information in statistical physics is based on the view of systems as large collections of interacting deterministic subsystems, in the early 20th century **quantum theory** extended the range of uncertainty to microscopic physics. Heisenberg's Uncertainty Principle expresses fundamental limits on the certainty with which *any* physical system can be described, irrespective of size or complexity. The deterministic states of the Newtonian framework dissolve into probabilistic meta-states, whose development can by described only by the stochastics of Schrödinger's state transition function. Wheeler [303] has described how informational concepts are gaining importance in explanations of quantum theory.

### 1.2.5 Algorithmic Information

With the advent of programmable machines, the concept of the **algorithm** came to prominence, and the general characteristics and properties of algorithms and machines which implement them were investigated. Measures of time and space complexity were identified as informational concepts, and such measures as **algorithmic depth** [11] were developed.

**Algorithmic complexity**, also sometimes called **algorithmic information**, was described independently by Kolmogorov [162] and Chaitin [29] as a measure of the "compressibility" of a symbol string by some compression algorithm. This quantity can in turn can be taken as the information content of a system relative to some other interpreting computational system.

### 1.2.6 Chaotic Information

Also with the advent of computing machines the approximations of dynamical systems without analytic solutions became more tractable and complete. As researchers began empirical exploration of the space of such systems, the special mathematical properties of these systems, first discovered by Poincare, was developed into the theory of **chaos**.

The development of chaotic dynamics in the 1980's added to the now rich field of information theories. In chaotic dynamical systems, while short term prediction of trajectories is completely possible and deterministic, errors in calculations and small differences in the specification of initial conditions do not damp out, but rather grow exponentially in model time. Thus while microscopically deterministic, at the macroscopic level their behavior is highly unpredictable, and it is common to resort to information theoretic descriptions at this level. The concept of **metric entropy** [258], also developed by Kolmogorov, is used in this context to measure the "spread" of the trajectories in the phase space.

Chaos provided a wide domain of differential systems which previously had no solution with a qualitative, macroscopic, stochastic description at a higher level of analysis. Thus chaotic information theory has had a major impact on physics. The first mathematical explorations of chaotic systems were performed on models of atmospheric convection [179]. A vast array of physical systems have proven to have excellent descriptions under chaotic dynamics, from some relatively simple physical systems like compound pendulums and dripping faucets to orbital dynamics and complex chemical and biophysical systems [10].

### 1.2.7 Computational Physics

One leading edge in computational information theory is the growth of "computational physics", which concerns itself specifically with the relation between computational information theory and physics. This discussion addresses many aspects of this relationship, including issues in algorithmic information theory and chaotic dynamics, but also about the nature of irreversibility in computation, the relation between measurement and computation, the thermodynamics of gravitational singularities, the complexity of physical systems, and the fundamental physical limits of computing machines [202, 307, 329].

### 1.2.8 Modeling and Simulation Technology

Finally, another very strong trend in the 1980's is the development of a number of computational techniques which are intended to model or simulate natural systems. Examples include neural network theory, which simulates neural systems; artificial intelligence, which simulates cognitive systems; genetic algorithms, which simulate genetic evolution [121]; and a variety of "artificial life" systems which simulate a variety of different organismal behaviors [172].

Most of the above technologies, while very popular, hold at best a metaphorical relationship to the systems they are intended to model. As discussed by Pattee [201] and Rosen [242], they suffer from the deficiency of being "mere" simulations, which cannot capture much of the causal structures of their object systems. Nevertheless, they are capable of very good mimicry, and certainly their *intention* is to maintain as close a relation as possible to the systems they model.

An example of a better technique which actually tries to provide a computer model of the causal relations in a complex biological systems is the work of William Powers [215, 216]. His hierarchical feedback control models are a striking blending of theory and simulation, and promise to provide a key link in the application of informational concepts to complex physical systems.

## 1.3 Information Science as Meta-Science

Unlike physical concepts like energy and mass, informational concepts like organization and entropy have a special dual nature, in that they relate to both objective and subjective phenomena and attributes. Therefore they can be applied not only to the *objects* of science, for example by considering a thermodynamic or biological system in informational terms, but also reflexively to the *processes* of science itself. Therefore the various branches of information science can be used reflexively to analyze the scientific process as a whole, resulting in the possibilities for both **unified science** and **meta-science**.

A prime example of this is in the application of statistical techniques to the problems of induction, evidence, inference, and theory construction, as typified by the classical work of Reichenbach [236] and Salmon [245]. Other examples include the work of Rescher [237, 238], who uses informational concepts to explore the limits to theory construction; and Klir [145] who has constructed a number of sophisticated multidimensional inductive modeling methodologies based on strong information and systems scientific principles.

One of the most interesting approaches in this spirit is the work of the **maximum entropy** school begun by Ed Jaynes with his classic paper on statistical physics [127]. In it he turned the traditional relation between thermodynamic and informational entropy on its head. Instead of seeing maximum statistical entropy as a derived *conclusion*, a reflection of the complete ignorance of the observer relative to an underlying "real" process, he regards it as an *assumption* to the process of deriving scientific laws. Given some boundary constraints, for example from the law of conservation of energy, and then performing an entropy maximization problem, the basic results of statistical physics can be directly derived. Jaynes and his followers [289, 269] as well as others [35, 137] have developed this **Maximum Entropy**

**Principle** (MEP) to apply to many problems in scientific inference.

All of the above programs can be ultimately regarded as projects in scientific **semiotics**, in which the processes of science are regarded as processes in symbolic systems. It is through the science of symbolic systems that information science approaches psychology in the abstract, where mental processes are regarded either microscopically as a parallel computational process in a neural system, or macroscopically as a system of representations maintained by an organism of its environment.

In any of these manifestations, the view of information science as metascience is inherently connected to the underlying natural scientific theories to which it is applied.

## 1.4 Generalized Information Theory

The final branch of information science to be discussed is that which will occupy us for the rest of this work, and which we will call, following Klir [151], **Generalized Information Theory**. GIT has many components, including **fuzzy sets and measures, uncertainty measures, evidence theory, random set theory, possibility theory**, and traditional **probability theory** as a special case. These specific aspects will be examined in detail in Chap. 2, with only some of their overall characteristics discussed here.

The components of GIT are all mutually interrelated, and are all related to probability theory and classical information theory. None of them are completely radical departures, but rather are derived by relaxing fundamental principles of a traditional theory, for example the excluded middle (fuzzy sets), additive measures (fuzzy measures), additive normalization (possibility theory), point-valued random variables (random sets), or point-valued probability distributions (evidence theory).

### 1.4.1 GIT and Science

The explosive development of physical science in this century is marked with a concomitant growth in information science, to the point that now the relation between them is highly complex and intimate. Each major branch of the information sciences, from information theory to computational theory to metascience, has developed in close relation to natural science. This follows the pattern in many branches of mathematics, where scientific and mathematical development are mutually reinforcing.

GIT was founded in the mid-1960's, and has achieved a great fruition in the 1990's. But it is surprising that these new mathematical theories generally do *not* follow in the tradition of information science, and are *not* (generally) closely related to any naturalistic scientific theories or applications.

Indeed, we now stand at a time of a flowering of mathematical information theories, an expanding universe of formalisms which provide generalizations of generalizations along all dimensions of the classical theories. This movement takes place amidst great excitement about the potential to integrate these methods into existing and new techniques in a variety of engineering disciplines. Yet at the same time there is very little attention being paid to the potential significance of the new information theories for science and meta-science.

As we approach the new millennium we face an historically unique situation: today there are aspects of information science which have exceeded natural science to the extent that it is unknown what, if any, relation there might be between them and the natural laws which govern the world. While this situation is not unusual for mathematics in general, where there need be no necessary relation between a new mathematical and a new physical theory, it is unusual for information science, growing up as it has closely tied to scientific theory.

That is not to say that these theories lack applications of any kind. On the contrary, there is currently a great deal of interest, verging on faddishness, in GIT methods. But virtually none of this interest or these applications can be described as being related to natural science in the same sense that traditional information theory has been.

### 1.4.2 Applications in "Informational Engineering"

The vast majority of the applications of GIT are in what could be described as **informational engineering**, conceived of as the application of computer and electronic technology to the management of human technological systems. Examples include application to control system technology, decision support systems, pattern matching systems, expert systems, approximate reasoning systems, and a variety of other "knowledge based" applications.

### 1.4.3 Applications in Psychology

A common factor of the above applications is the emphasis that they place on the management of human-controlled or human-modeled systems: a knowledge-based control system is intended to replace a human operator; an expert system is intended to replace a human expert; a pattern-matching system is intended to emulate human perception; and an approximate reasoning system is intended to emulate human reasoning.

More strongly, many branches of GIT were founded specifically for the purpose of modeling human psychology. This is certainly the case for two of its major branches, those of fuzzy sets and evidence theory. In the founding paper on fuzzy set theory, Zadeh says

> Clearly, the "class of all real numbers which are much greater than 1", or "the class of beautiful women", or "the class of tall men", do not constitute classes or sets in the usual mathematical sense of these terms. Yet, the fact remains that such imprecisely defined "classes" play an important role in human thinking, particularly in the domains of pattern recognition, communication of information, and abstraction. [321]

Or in the paper introducing the concept of the "fuzzy restriction", which will later lead to the possibility distribution, he says

> A common thread that runs through most of the applications of the theory of fuzzy sets relates to the concept of a *fuzzy restriction* — that is, a fuzzy relation which acts as an elastic constraint on the values that may be assigned to a variable. Such restrictions appear to play an important role in human cognition, especially in situations involving concept formation, pattern recognition, and decision-making in fuzzy or uncertain environments. [323]

At no point does Zadeh ask what an "elastic constraint" might be in the context of *physical* systems, but merely assumes that it results from the vagaries of human judgment.

As an example of the fulfillment of this principle, in the majority of fuzzy systems applications the sole representation used is that of the "linguistic variable", intended to model a subjective term as used by a human.

Similarly, Shafer [261] developed his theory of belief functions, later to grow into the Dempster-Shafer theory of evidence, specifically to model human subjects tendencies to report ranges of probability estimates.

### 1.4.4 Lack of Objective Semantics

In light of this historical development of GIT, it is not surprising that the semantics of fuzzy theory and possibility theory have been based almost exclusively on the subjective opinions of people. There is a deep assumption in GIT that, in comparison with traditional information theory, the values of fuzzy set membership grades, or possibilities, or beliefs, are inherently meant to model the uncertainty of a human subject. This assumption may be implicit, but it is also frequently stated explicitly, and with little or no justification, for example by Kandel, "Probability is an *objective* characteristic. The membership grade is subjective [132]."

Of course classical information theory also has applications in informational engineering, and that is not an indictment of it. And the subjectivist school of probability semantics has made an important contribution to probability theory. My criticism of GIT is not that it *includes* these applications and these semantics, but rather that it is *limited* to them in a way in which traditional information theory is not.

Whereas traditional information theory is deeply rooted in physics and biology, and has *grown* to include human psychology, GIT was *founded* entirely in the domain of human psychology, and with few exceptions has not grown *out of* it. It does not seem ludicrous to try to find an objective semantics for GIT methods. It rather seems that a complacent research community has been satisfied with ceding the "real world" to the probabilists for thirty years.

## 1.5 Towards an Objective Possibility Theory

This condition is what the present work is attempting to alleviate, by defining possibility as an independent form for the representation of uncertainty, and moving towards an objective, empirical semantics of possibility through the development of **possibilistic measurement procedures** and **possibilistic models** of physical systems.

In Chap. 2 the fundamental existing results and some new results in mathematical possibility theory will be introduced in the context of the various components of GIT and their relationships. Rather than the traditional approach of grounding possibility theory in the theory of fuzzy subsets, we will approach possibility theory from the perspective of random subset theory and distribution functions of fuzzy measures. It will then be argued that possibility theory has a legitimate identity as a form of information theory distinct from both probability and fuzzy sets.

The semantics of possibility theory are discussed in Chap. 3. After arguing that any legitimate natural semantics of possibility must rely on the dual procedures of measurement and interpretation, the roots of possibilistic semantics in philosophy, natural language, modal logic, and especially probability theory will be discussed. After examining the semantic consequences which flow from the formalism of mathematical possibility theory itself, the traditional semantics of possibility theory will be critiqued.

Chap. 4 is dedicated to the development of a variety of objective measurement procedures for possibility distributions. These are based on measurement methods which yield subsets, not points, either directly from a measuring instrument or a set of measuring instruments, or indirectly through the manipulation of point data streams. The results of these procedures compare favorably with the fuzzy number forms of possibility distributions used in many applications.

Possibilistic processes and automata are introduced in Chap. 5. General processes are based on semirings, and the relation between semirings and t-norm and t-conorm pairs is disambiguated. Familiar classes of processes include the deterministic, nondeterministic, and stochastic cases. Possibilistic processes are a new class. They are placed in the context of general automata and the other classes of processes, and it is argued that they are the legitimate generalization of nondeterministic processes, in a way that stochastic processes are not. Conditional possibility measures and distributions, used in possibilistic processes, are also available on semirings. Possibilistic automata and other possibilistic machines are also discussed, including possibilistic networks, Monte Carlo methods, and Markov processes.

Chap. 6 presents the architecture for an implementation of possibilistic methods in an object-oriented, C++ environment in the context of the Computer-Aided Systems Theory (CAST) research program. Implementation of both fundamental and supplementary GIT methods is proposed, as well as links to other aspects of GIT and other CAST implementations. Such a system is necessary both for the implementation of possibilistic models, as well as for the empirical exploration of the properties of possibilistic systems and processes.

In its integration of interval-based and uncertainty distribution methods, possibility theory promises to provide an important new set of methods for the new movement in qualitative modeling (QM). Recently such models have been applied to systems which diagnose the possible faults in complex systems like spacecraft. So finally Chap. 7 considers the promising potential for the application of possibilistic modeling methods to the qualitative model-based diagnosis and trend analysis of spacecraft systems.

# Chapter 2

# Possibility Theory in Generalized Information Theory

Let $X = X$.

— *Laurie Anderson*

At the most general level, GIT consists of the fields of **fuzzy sets**, **fuzzy measures**, and **uncertainty measures**. Both **probability theory** and **possibility theory** are related in that their measures are fuzzy measures and their distributions are fuzzy sets. **Evidence theory** and **random set theory** are closely related fields in which both fuzzy measures, as set functions, and fuzzy sets, as point functions, are unified.

In this chapter some new and some established results of possibility theory in relation to the branches of GIT will be presented. In doing so the various components of GIT and the mathematical notation and concepts used will be formally introduced and interrelated. Much of this material is drawn from the published literature,[1] and these results and definitions will be interwoven with original results and novel notation. A number of the results presented are not claimed to be original, but are presented for the sake of completeness. Part of the purpose is to present this material in our specific mathematical context, in which possibility theory is based on consistent random sets. Citations to the literature will be offered wherever a

[1]The reader is referred to some of the standard texts on these subjects: for fuzzy sets and systems to Zimmerman [328] and Dubois and Prade [55]; for fuzzy measures to Wang and Klir [299]; for evidence theory to Shafer [261]; for random set theory to Kendall [142] and Goodman and Nguyen [104]; for possibility theory to Dubois and Prade [64]; and for uncertainty measures and GIT as a whole to Klir [154] and Klir and Folger [155].

result is drawn explicitly from an external source, even if an alternate definition or proof is provided.

But a number of new results in possibility theory *are* presented, in particular the relation between the distributions of fuzzy measures and structural and numerical aggregation functions on random sets; and well-justified possibilistic normalization procedures in the context of random sets.[2]

Finally, it is argued that contrary to the standard interpretations, possibility theory is a new form of information theory, distinct from, but related to, both probability and fuzzy sets. The traditional ideas from fuzzy systems theory that there is a special equivalence between fuzzy sets and possibility distributions, that possibilistic concepts are inherently more appropriate for fuzzy theory than probabilistic concepts are, and that fuzzy theory is inherently possibilistic and non-probabilistic, are rejected. Also, even though possibility theory is in a number of ways weaker than probability, it is nevertheless logically independent of it.

It is assumed that the reader is conversant in the various aspects of traditional information theory on which there is a vast literature: probability and statistics, classical measure theory, classical information theory, and some rudimentary abstract algebra.

Throughout the following a finite universe of discourse $\Omega = \{\omega_i\}, 1 \leq i \leq n < \infty$ will generally be assumed. Occasionally $\Omega = \{\omega\}$ will be used when referring to a continuous universe or a universe with unspecified cardinality.

## 2.1 Algebraic Preliminaries

In this section, some of the mathematical structures used by many components of GIT are defined. While these structures are usually based on the unit interval $[0,1]$ with the total order $\leq$, there are generalizations to full lattices with partial orders. Both cases will be presented here.

Let $\langle \mathcal{L}, \leq \rangle$ be a complete lattice with $0, 1 \in \mathcal{L}$ as the global supremum and infimum. $\vee$ and $\wedge$ will denote either the join or meet in $\mathcal{L}$ or the maximum and minimum operators on $[0,1]$ as appropriate.

### 2.1.1 Complements

**Definition 2.1 (Complement (Lattice))** [38] A function $\varphi : \mathcal{L} \mapsto \mathcal{L}$ is a **complement** function if:

[2] A complete list of the original contributions of this dissertation is provided in Appendix A.

- **Boundary Condition:** $\varphi(0) = 1$.
- **Monotonicity:** $\forall x, y \in \mathcal{L}, x \leq y \rightarrow \varphi(x) \leq \varphi(y)$.
- **Involution:** $\forall x \in \mathcal{L}, \varphi(\varphi(x)) = x$.

**Corollary 2.2** $\varphi(1) = 0$.

**Proof:** $\varphi(\varphi(0)) = 0$ and $\varphi(\varphi(0)) = \varphi(1)$, so $\varphi(1) = 0$. ∎

**Definition 2.3 (Complement (Standard))** [56] A complement function $\varphi$ on $[0, 1]$ is a **standard complement** if $\varphi$ is continuous.

Almost always $\varphi(x) = 1 - x$.

### 2.1.2 Norms and Conorms

Norm and conorm operators are dual, and their definitions will be made "in parallel".

**Definition 2.4 (Lattice Norm (Conorm))** [38] The function $\sqcap: \mathcal{L}^2 \mapsto \mathcal{L}$ (resp. $\sqcup: \mathcal{L}^2 \mapsto \mathcal{L}$) is a **norm** (**conorm**) if

- $\langle \mathcal{L}, \sqcap, 1 \rangle$ (resp. $\langle \mathcal{L}, \sqcup, 0 \rangle$) is an Abelian monoid (commutative and associative operator with identity 1 (resp. 0));
- $\sqcap$ and $\sqcup$ are monotonic:

$$\forall, x_1, x_2, y_1, y_2 \in \mathcal{L}, \quad x_1 \leq x_2, y_1 \leq y_2 \quad \rightarrow \quad x_1 \sqcap y_1 \leq x_2 \sqcap y_2, \quad x_1 \sqcup y_1 \leq x_2 \sqcup y_2.$$

**Corollary 2.5** 0 (resp. 1) is a zero of $\sqcap$ ($\sqcup$), that is

$$\forall x \in \mathcal{L}, \quad x \sqcap 0 = 0 \sqcap x = 0 \text{ and } x \sqcup 1 = 1 \sqcup x = 1.$$

**Proof:** Since $0 \leq 0$ and $\forall x \in \mathcal{L}, x \leq 1$, then

$$\forall x \in \mathcal{L}, x \sqcap 0 \leq 1 \sqcap 0 \leq 0.$$

Since $\inf \mathcal{L} = 0$, therefore $x \sqcap 0 = 0$. The other result for $\sqcap$ follows from commutivity. The results for $\sqcup$ follow analogously. ∎

**Corollary 2.6** $\forall x, y, z \in \mathcal{L}, y \leq z \rightarrow x \sqcap z \leq x \sqcap w$.

**Proof:** Follows from $x \leq x$ and monotonicity. ∎

$\sqcap$ and $\sqcup$ are dual, related by DeMorgan's property under $\varphi$.

| | Idempotent | | Monotonic | | Nilpotent | | "Crisp" |
|---|---|---|---|---|---|---|---|
| $\sqcap$ | $x \wedge y$ | $\geq$ | $x \times y$ | $\geq$ | $x \sqcap_m y := 0 \vee (x + y - 1)$ | $\geq$ | $x \sqcap_w y := \lfloor x \rfloor \lfloor y \rfloor$ |
| $\sqcup$ | $x \vee y$ | $\leq$ | $x \star y := x + y - xy$ | $\leq$ | $x \sqcup_m y := 1 \wedge (x + y)$ | $\leq$ | $x \sqcup_w y := \lceil x \rceil \lceil y \rceil$ |

Table 2.1: Common $\langle \sqcap, \sqcup \rangle$ pairs on $[0, 1]$, and their relation.

**Proposition 2.7** [38] Assume $\sqcap, \sqcup$, and for some $\varphi$ define the DeMorgan transforms

$$x \sqcap^\varphi y := \varphi(\varphi(x) \sqcup \varphi(y)), \qquad x \sqcup^\varphi y := \varphi(\varphi(x) \sqcap \varphi(y)).$$

Then there exists a $\varphi$ such that $\sqcap^\varphi$ is a conorm, $\sqcup^\varphi$ is a norm, and

$$(\sqcap^\varphi)^\varphi = \sqcap, \qquad (\sqcup^\varphi)^\varphi = \sqcup.$$

$\langle \vee, \wedge \rangle$ is a norm/conorm pair, and $\langle \sqcap_w, \sqcup_w \rangle$ is the norm/conorm pair defined by

$$x \sqcap_w y := \begin{cases} x, & y = 1 \\ y, & x = 1 \\ 0, & \text{otherwise} \end{cases}, \qquad x \sqcup_w y := \begin{cases} x, & y = 0 \\ y, & x = 0 \\ 1, & \text{otherwise} \end{cases}.$$

All of the above conditions hold when $\langle \mathcal{L}, \leq \rangle = \langle [0, 1], \leq \rangle$. Table 2.1 summarizes some of the prominent norm/conorm pairs available on $[0, 1]$, where $\lfloor x \rfloor$ and $\lceil x \rceil$ are the least and greatest integers near $x$ respectively.

In general,

$$x \wedge y \geq x \sqcap y \geq x \sqcap_w y \qquad x \vee y \leq x \sqcup x \sqcup_w y.$$

## 2.2 Possibility Theory

The concept of "possibility" has a long history in philosophy. Some of these ideas have been expressed in a formal theory beginning with modal logic (see Sec. 3.2). But modal logic expresses possibility as a *crisp* concept, that is, representing either completely possible or completely impossible events. Mathematical possibility theory differs from modal logic in that it is a formal theory which represents possibility which admits to *degrees* between and including complete possibility and impossibility, and usually expressed on the unit interval.

Possibility theory was first formalized and axiomatized by Shackle [260]. It was later introduced again by Zadeh [325], who related it strictly to his fuzzy set theory. Possibility measures and distributions also arise in the context of fuzzy measures and random sets.

There have been a number of recent efforts to axiomatize possibility on the basis of qualitative relations by Dubois and Prade [52, 66], the semantics of betting by Giles [98], measurement theory by Yager [311], abstract algebra by Yager [318], and lattice-theoretic based measure theory by Cooman et al. [38]. A very general axiomatization of the core tenets will first be presented (joint and conditional possibility will not be discussed until Chap. 5), followed by the specialization to the more standard version used in the general literature.

### 2.2.1 Generalized Possibility

The following is a very general axiomatization combining the work of Yager [318] and Cooman et al. [38].

Let $\mathcal{B}$ be a finite Boolean algebra with $\mathbf{0}, \mathbf{1} \in \mathcal{B}$ as the global supremum and infimum. Let $\vee, \wedge$ be the join and meet in both $\mathcal{B}$ and $\mathcal{L}$ without ambiguity, and for $b \in \mathcal{B}$, let $\overline{b} \in \mathcal{B}$ be the complement of $b$ in $\mathcal{B}$.

**Definition 2.8 (Possibility Measure (General))** A **possibility measure** is a function $\Pi: \mathcal{B} \mapsto \mathcal{L}$, where

- $\Pi(\mathbf{0}) = 0$.
- $\forall \{b_j\} \subseteq \mathcal{B}$ with $j \in J$ an arbitrary index,

$$\Pi\left(\bigvee_{j \in J} b_j\right) := \bigvee_{j \in J} \Pi(b_j).$$

**Definition 2.9 (Necessity Measure (General))** Assume $b_j$ as above. A **necessity measure** is a function $\eta: \mathcal{B} \mapsto \mathcal{L}$ where

- $\eta(\mathbf{1}) = 1$.
- $\forall \{b_j\} \subseteq \mathcal{B}$ with $j \in J$ an arbitrary index,

$$\eta\left(\bigwedge_{j \in J} b_j\right) := \bigwedge_{j \in J} \eta(b_j).$$

**Proposition 2.10** [38] Let $\Pi^\varphi: \mathcal{B} \mapsto \mathcal{L}$ and $\eta^\varphi: \mathcal{B} \mapsto \mathcal{L}$ with

$$\Pi^\varphi(b) = \varphi(\Pi(\overline{b})), \qquad \eta^\varphi(b) = \varphi(\eta(\overline{b})).$$

Then $\Pi^\varphi$ is a necessity measure dual to $\Pi$, and $\eta^\varphi$ is a possibility measure dual to $\eta$.

Results for possibility generally hold in the dual for necessity. Therefore necessity measures will be used sporadically in the sequel as required.

**Definition 2.11 (Normalization (General))** $\Pi$ is **normal** if $\Pi(\mathbf{1}) = 1$.

**Definition 2.12 (Atoms)** [318] $a \in \mathcal{B}, a \neq \mathbf{0}$ is an **atom** of $\mathcal{B}$ if

$$\forall b \in \mathcal{B}, a \wedge b = a \text{ or } a \wedge b = \mathbf{0}.$$

Let $\mathcal{M}(\mathcal{B}) := \{a_i\}$ be the set of all atoms of $\mathcal{B}$. $\forall b \in \mathcal{B}$, let

$$\mathcal{M}(b) := \left\{ a_i \in \mathcal{M}(\mathcal{B}) : b = \bigvee_i a_i \right\}.$$

Denote $a_i, b$ as a generic atom and element of $\mathcal{B}$ respectively.

**Definition 2.13 (Possibility Distribution (General))** A **possibility distribution** is a function $\pi \colon \mathcal{M}(\mathcal{B}) \mapsto \mathcal{L}$ with

$$\pi(a_i) := \Pi(a_i).$$

**Corollary 2.14**

$$\Pi(b) = \bigvee_{a_i \in \mathcal{M}(b)} \pi(a_i).$$

**Proof:** Trivial from the definition of the possibility measure (2.8) and lattice atoms (2.12). ■

**Definition 2.15 (Possibility Distribution Core)** The **core** of a possibility distribution is $\mathbf{C}(\pi) := \{a_i : \pi(a_i) = 1\}$.

**Definition 2.16 (Possibility Distribution Focus)** If $\exists a^* \in \mathcal{M}(\mathcal{B})$ with $a^* \in \mathbf{C}(\pi)$, then $a^*$ is a **focus** of $\pi$.

**Corollary 2.17 (Existence of a Focus)** [38] If $\Pi$ is normal, then $\exists a^* \in \mathcal{M}(\mathcal{B}), \pi(a^*) = 1$.

**Proof:** From normalization (2.11),

$$\Pi(\mathbf{1}) = \bigvee_{a_i \in \mathcal{M}(\mathbf{1})} \pi(a_i) = \bigvee_{a_i \in \mathcal{M}(\mathcal{B})} \pi(a_i) = 1.$$

Since 1 is the supremum of $\mathcal{L}$, if $\nexists a^*, \pi(a^*) = 1$, then $\bigvee_{a_i} \pi(a_i) \neq 1$. Therefore $\exists a^* \in \mathbf{C}(\pi)$. ■

**Corollary 2.18** If $\Pi$ is normal then $\Pi(b) \vee \Pi(\bar{b}) = 1$.

**Proof:** $\Pi(b) \vee \Pi(\bar{b}) = \Pi(b \vee \bar{b}) = \Pi(\mathbf{1}) = 1$. ■

### 2.2.2 Specializations

Possibility theory has almost always been developed in the context where $\mathcal{L} = [0, 1]$ and $\leq$ is the standard total order on the reals. Cooman et al. [38] were the first to extend $[0, 1]$ to a full lattice. Dubois, Lang, and Prade [53] have also considered possibility measures valued on the lattice of fuzzy sets. Cooman et al. are quite clear that some of the more standard results of possibility theory only hold when $\leq$ is a complete, and not a partial, order.

**Proposition 2.19** [38] Let $\mathcal{L} = [0, 1]$, and let $\Pi$ and $\eta$ be dual for some complement $\varphi$. Then $\Pi(b) < 1 \to \eta(b) = 0$ and $\eta(b) > 0 \to \Pi(b) = 1$.

**Proposition 2.20** [38] Assume $\mathcal{L}, \Pi, \eta$ as above. Then $\Pi(b) \geq \eta(b)$.

Possibility theory has also almost always been developed in the context where $\mathcal{B} = 2^\Omega$ is the power set of some (usually finite) set $\Omega$. Then the operators $\vee, \wedge, \leq$ in $\mathcal{B}$ become $\cup, \cap, \subseteq$ in $2^\Omega$, and

$$\mathbf{1} = \Omega, \qquad \mathbf{0} = \emptyset.$$

$$\mathcal{M}(\mathcal{B}) = \mathcal{M}(2^\Omega) = \{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}\}, \qquad \forall A \subseteq \Omega, \mathcal{M}(A) = \{\{\omega_i\} : \omega_i \in A\}.$$

When both specializations of $\mathcal{B}$ to $2^\Omega$ and $\mathcal{L}$ to $[0, 1]$ are made, then the following standard definitions are achieved.

**Definition 2.21 (Possibility Measure (Standard))** A possibility measure is a function $\Pi \colon 2^\Omega \mapsto [0, 1]$ with finite $\Omega$ where

- $\Pi(\emptyset) = 0$.
- $\forall \{A_j\} \subseteq 2^\Omega$ with $j \in J$ an arbitrary index,

$$\Pi\left(\bigcup_{j \in J} A_j\right) := \bigvee_{j \in J} \Pi(A_j).$$

**Definition 2.22 (Normalization (Standard))** $\Pi$ is normal if $\Pi(\Omega) = 1$.

**Definition 2.23 (Possibility Distribution (Standard))** A possibility distribution is a function $\pi \colon \Omega \mapsto [0, 1]$ with

$$\pi(\omega) := \Pi(\{\omega\}).$$

**Corollary 2.24**

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi(\omega_i).$$

**Proof:**

$$\Pi(A) = \bigvee_{\{\omega_i\} \in \mathcal{M}(A)} \Pi(\{\omega_i\}) = \bigvee_{\omega_i \in A} \pi(\omega_i).$$

■

**Definition 2.25 (Crisp Possibility Distribution)** A possibility distribution is **crisp** if $\forall \omega \in \Omega, \pi(\omega) \in \{0,1\}$.

**Corollary 2.26 (Possibilistic Normalization)** If $\Pi$ is normal, then $\exists \omega^*, \pi(\omega^*) = 1$.

**Proof:** Directly from the general result (2.17). ■

## 2.3 Fuzzy Measures

Fuzzy measures were introduced by Sugeno [276,277] in order to generalize classical measures [115], and were extended by Wang [298] and Wang and Klir [299].

**Definition 2.27 (Fuzzy Measure and Fuzzy Measure Space)** [299] A **fuzzy measurable space** is $\langle \Omega, \Sigma \rangle$, where $\Sigma \subseteq 2^\Omega$ is a sigma-field of $\Omega$. A **fuzzy measure** on a fuzzy measurable space is a function $\nu \colon \Sigma \mapsto [0, \infty]$, where:

- If $\emptyset \in \Sigma$ then $\nu(\emptyset) = 0$,
- $\forall A, B \in \Sigma$, if $A \subseteq B$ then $\nu(A) \leq \nu(B)$.

Then the fuzzy measure space is $\langle \Omega, \Sigma, \nu \rangle$.

It is common to let $\nu \colon 2^\Omega \mapsto [0,1]$, which will be assumed in the sequel.

**Proposition 2.28 (Fuzzy Measure Limits)** [299] Assume a fuzzy measure space $\left\langle \Omega, 2^\Omega, \nu \right\rangle$ to $[0,1]$. Then

$$\nu(A \cup B) \geq \nu(A) \vee \nu(B), \qquad \nu(A \cap B) \leq \nu(A) \wedge \nu(B).$$

**Corollary 2.29** A probability measure Pr on $\left\langle \Omega, 2^\Omega \right\rangle$ is a fuzzy measure on $2^\Omega$ to $[0,1]$.

**Proof:** Let $A \subseteq B$. Then $B = A \cup (B - A)$, and $A \cap (B - A) = \emptyset$, so that $\Pr(A \cup B) = \Pr(B) = \Pr(A) + \Pr(B - A)$. Therefore $\Pr(B) \geq \Pr(A)$. ■

**Corollary 2.30** If $\Omega$ is finite then $\Pi$ is a fuzzy measure on $2^\Omega$ to $[0,1]$.

**Proof:** If $A \subseteq B$ then $A \cup B = B$, so that $\Pi(A \cup B) = \Pi(A) \vee \Pi(B) = \Pi(B)$, and $\Pi(A) \leq \Pi(B)$. ■

Puri and Ralescu [224] have shown that when $\Omega$ is not finite, then if $\Pi$ is fully continuous then it in fact is *not* a full fuzzy measure, although Wang and Klir [299, p. 63] have shown that they are lower semi-continuous fuzzy measures.

$\nu$ is a set function on $2^\Omega$, and thus the size of its domain grows exponentially with $|\Omega|$. The purpose of a **distribution** of a fuzzy measure is to recover knowledge of $\nu$ from values only on elements of $\Omega$, which of course grows only linearly with $|\Omega|$.

**Definition 2.31 (Distribution of a Fuzzy Measure)** Given a fuzzy measure $\left\langle \Omega, 2^\Omega, \nu \right\rangle$ to $[0,1]$ with $\Omega$ finite and $|\Omega| = n$, then the function

$$q_\nu \colon \Omega \mapsto [0,1], \qquad q_\nu(\omega) := \nu(\{\omega\}),$$

is a **distribution** of $\nu$ if there exists a distribution operator function $\oplus \colon [0,1]^2 \mapsto [0,1]$ where

- $\langle [0,1], \oplus, 0 \rangle$ is an Abelian monoid ($\oplus$ is a commutative, associative, operator with identity 0).
- In operator notation,

$$\forall A \subseteq \Omega, \quad \bigoplus_{\omega \in A} q_\nu(\omega) = \nu(A). \tag{2.32}$$

## 2.4 Fuzzy Sets

A **crisp subset** $F \subseteq \Omega$ is denoted by its **characteristic function**

$$\chi_F \colon \Omega \mapsto \{0,1\}, \qquad \chi_F(\omega) = \begin{cases} 1, & \omega \in F \\ 0, & \omega \notin F \end{cases}.$$

Fuzzy sets result from a direct generalization.

**Definition 2.33 (Fuzzy Set)** [320] A **membership function** is a function $\mu \colon \Omega \mapsto [0,1]$. A **fuzzy subset** on $\Omega$, denoted $\widetilde{F} \mathrel{\tilde{\subseteq}} \Omega$, implies the existence of a membership function $\mu_{\widetilde{F}} \colon \Omega \mapsto [0,1]$ such that

$$\mu_{\widetilde{F}}(\omega) = \begin{cases} 1, & \omega \in \widetilde{F} \\ 0, & \omega \notin \widetilde{F} \end{cases},$$

and $\mu_{\widetilde{F}}(\omega) \in (0,1)$ means that $\omega \in \widetilde{F}$ "to the extent of" $\mu_{\widetilde{F}}(\omega)$.

Goguen [100] has extended fuzzy sets to lattice-valued, or L-fuzzy, sets similar to Cooman et al.'s generalization of possibility measures to a lattice.

Since $\{0,1\} \subseteq [0,1]$, therefore all crisp sets are fuzzy sets. The **extension principle** states, in part, that any fuzzifying generalization must be consistent with the classical, crisp special cases. The extension principle is not determinative, since many different fuzzifications may have the same values in the crisp cases. Thus the crisp cases form only a partial constraint on the admissible fuzzy generalizations.

**Proposition 2.34** A fuzzy measure $\nu$ to $[0,1]$ with distribution $q_\nu$ induces the fuzzy sets denoted

$$\widetilde{\nu} \mathrel{\tilde{\subseteq}} 2^\Omega, \qquad \mu_{\widetilde{\nu}}(A) := \nu(A),$$

$$\widetilde{q}_\nu \mathrel{\tilde{\subseteq}} \Omega, \qquad \mu_{\widetilde{q}_\nu}(\omega) := q(\omega).$$

Note that $\widetilde{\nu}$ is a fuzzy subset of $2^\Omega$, not $\Omega$.

### 2.4.1 The Fuzzy Power Set

A very useful representation for fuzzy sets is offered by Kosko [164, 165]. First, consider a crisp subset $F \subseteq \Omega$ represented as a set of tuples $\{\langle \omega_i, b_i \rangle\}, 1 \leq i \leq n$ where $b_i := \chi_F(\omega_i) \in \{0,1\}$. When the ordering of the $\omega_i$ are fixed, then they can be assumed, and the representation $F$ can be replaced by the simple bit vector, or bit string $\vec{F} := \langle b_i \rangle$.

$\vec{F}$ can be regarded as a vertex of the boolean hypercube of dimension $n$. Denote this lattice as $2^\Omega$, which is equivalently the **power set** of $\Omega$

$$2^\Omega = \{F \subseteq \Omega\}, \qquad \forall F \subseteq \Omega, F \in 2^\Omega.$$

This approach generalizes naturally to fuzzy sets. First represent a fuzzy subset $\widetilde{F} \mathrel{\tilde{\subseteq}} \Omega$ as a set of tuples $\{\langle \omega_i, f_i \rangle\}, 1 \leq i \leq n$, where $f_i := \mu_{\widetilde{F}}(\omega_i) \in [0,1]$. As above, when the ordering of the $\omega_i$ are assumed, then the vector representation $\vec{\widetilde{F}} := \langle f_i \rangle$ is arrived at. Kosko calls each $f_i$ a "fit", for "fuzzy digit", to map with "bit" for "binary digit".

The fuzzy set $\widetilde{F}$ is thus represented not as a point on a boolean $2^n$-lattice, but a point *inside* the *unit* hypercube of dimension $n$, called the **fuzzy power set** of $\Omega$ denoted

$$[0,1]^\Omega := \{\widetilde{F} \mathrel{\tilde{\subseteq}} \Omega\}, \qquad \forall \widetilde{F} \mathrel{\tilde{\subseteq}} \Omega, \widetilde{F} \in [0,1]^\Omega.$$

Of course $2^\Omega \subseteq [0,1]^\Omega$, in keeping with the extension principle.

An example is shown in Fig. 2.1 for $\Omega = \{x, y\}$, so that $2^\Omega = \{0,1\}^2 \subseteq [0,1]^\Omega = [0,1]^2$.

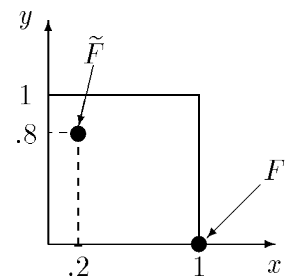


Figure 2.1: A crisp set $F = \{x\}$ and a fuzzy set $\widetilde{F} = \{\langle x, .2\rangle, \langle y, .8\rangle\}$.

### 2.4.2 Fuzzy Set Concepts

The following concepts related to fuzzy theory will also be used. Unless otherwise noted, see [155].

**Definition 2.35 (Operators)** Let $\widetilde{F}, \widetilde{G} \mathrel{\widetilde{\subseteq}} \Omega$ and $\sqcap, \sqcup$ be dual in $\varphi$. Then

$$\mu_{\widetilde{F}\cup\widetilde{G}} := \mu_{\widetilde{F}} \sqcup \mu_{\widetilde{G}}, \qquad \mu_{\widetilde{F}\cap\widetilde{G}} := \mu_{\widetilde{F}} \sqcap \mu_{\widetilde{G}}, \qquad \mu_{\overline{\widetilde{F}}} := \varphi\left(\mu_{\widetilde{F}}\right).$$

**Definition 2.36 (Inclusion)** Given $\widetilde{F}, \widetilde{G} \mathrel{\widetilde{\subseteq}} \Omega$, then $\widetilde{F} \subseteq \widetilde{G}$ if $\mu_{\widetilde{F}} \leq \mu_{\widetilde{G}}$.

**Definition 2.37 (Alpha Cut)** For $\alpha \in [0, 1]$,

$$\widetilde{F}_\alpha := \{\omega_i : \mu_{\widetilde{F}}(\omega_i) \geq \alpha\} \subseteq \Omega.$$

Each fuzzy set $\widetilde{F} \subseteq \Omega$ can be represented as a combination of its $\alpha$-cuts, each a crisp subset of $\Omega$, weighted by its $\alpha$ value. Thus

$$\mu_{\widetilde{F}}(\omega_i) = \max_{\alpha\in[0,1]} \omega_i \in \widetilde{F}_\alpha.$$

**Definition 2.38 (Level Set)** The set of distinct membership grades present in the fuzzy set:

$$\Lambda(\widetilde{F}) := \{\alpha \in [0, 1] : \exists\omega_i, \mu_{\widetilde{F}}(\omega_i) = \alpha\} \subseteq [0, 1].$$

$\Lambda(\widetilde{F})$ is the set produced by eliminating duplicates from the fit string representation $\vec{\widetilde{F}}$.

**Definition 2.39 (Support)** The crisp subset on which a fuzzy set has some membership:

$$\mathbf{U}(\widetilde{F}) := \{\omega_i : \mu_{\widetilde{F}}(\omega_i) > 0\} \subseteq \Omega.$$

**Definition 2.40 (Core)** The crisp subset on which a fuzzy set has complete membership:

$$\mathbf{C}(\widetilde{F}) := \widetilde{F}_1 \subseteq \Omega.$$

**Definition 2.41 (Fuzzy Set Normalization)** A normal fuzzy set has unity membership on at least one $\omega_i$, so that $\mathbf{C}(\widetilde{F}) \neq \emptyset$.

**Definition 2.42 (Fuzzy Relation)** A fuzzy subset $\widetilde{R} \mathrel{\widetilde{\subseteq}} \Omega_1 \times \Omega_2$ for some universes $\Omega_1, \Omega_2$.

**Definition 2.43 (Fuzzy Matrix)** Given a fuzzy relation $\widetilde{R} \mathrel{\widetilde{\subseteq}} \Omega_1 \times \Omega_2$, denote the **fuzzy matrix** as

$$R_{n_1 \times n_2} = [R_{ij}] := \left[\mu_{\widetilde{R}}(\omega_i, \omega_j)\right] \mathrel{\widetilde{\subseteq}} \Omega_1 \times \Omega_2,$$

where $n_k := |\Omega_k| < \infty$ for $k \in \{1, 2\}$.

### 2.4.3 Fuzzy Arithmetic

Fuzzy methods have been used to generalize many branches of mathematics (for example calculus), and fuzzy generalizations of arithmetic will be useful for us later on. All of the following are from Dubois and Prade [54,61]. See also Kaufmann and Gupta [139].

**Definition 2.44 (Fuzzy Interval)** A fuzzy subset of the real line $\widetilde{F} \mathrel{\widetilde{\subseteq}} \mathbb{R}$ where:

- $\widetilde{F}$ is normal.
- $\widetilde{F}$ is **convex:**

$$\forall\, x, y \in \mathbb{R}\,, \quad \forall\, z \in [x, y]\,, \quad \mu_{\widetilde{F}}(z) \geq \mu_{\widetilde{F}}(x) \wedge \mu_{\widetilde{F}}(y).$$

**Definition 2.45 (Fuzzy Number)** A fuzzy interval $\widetilde{F}$ where $\exists x \in \mathbb{R}, \mathbf{C}(\widetilde{F}) = \{x\}$.

Fuzzy intervals are direct generalizations of crisp intervals, and each fuzzy interval can be decomposed into its alpha cuts, a set of weighted crisp intervals.

**Proposition 2.46** If $\widetilde{F}$ is a fuzzy interval and $\mu_{\widetilde{F}}$ is upper semi-continuous, then

- $\forall \alpha \in (0, 1], \widetilde{F}_\alpha \subseteq \Omega$ is a closed interval.
- $\forall \alpha_1, \alpha_2 \in (0, 1], \alpha_1 \geq \alpha_2 \rightarrow \widetilde{F}_{\alpha_1} \subseteq \widetilde{F}_{\alpha_2}$.

**Proposition 2.47** A crisp interval $[a, b] \subseteq \mathbb{R}$ is a fuzzy interval $\widetilde{F}$ with

$$\mu_{\widetilde{F}}(x) = \begin{cases} 1, & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}, \qquad \forall \alpha \in (0, 1], \quad \widetilde{F}_\alpha = [a, b].$$

**Definition 2.48 (Unary Fuzzy Arithmetic Operators)** Let $f\colon \mathbb{R} \mapsto \mathbb{R}$ be an injective unary operator (for example, $f(x) = -x$ or $f(x) = e^x$) and $\widetilde{F}$ be a fuzzy interval. Then $\forall x \in \mathbb{R}$

$$\mu_{f(\widetilde{F})}(x) = \begin{cases} \mu_{\widetilde{F}}(f^{-1}(x)), & f^{-1}(x) \neq \emptyset \\ 0, & f^{-1}(x) = \emptyset \end{cases}.$$

**Definition 2.49 (Binary Fuzzy Arithmetic Operators)** Let $*$ be a binary arithmetic operator, for example $* \in \{+, -, \times, \div\}$, and $\widetilde{F}_1, \widetilde{F}_2$ be two fuzzy intervals. Then $\forall x \in \mathbb{R}$

$$\mu_{\widetilde{F}_1 * \widetilde{F}_2}(x) := \bigvee_{y*z=x} \mu_{\widetilde{F}_1}(y) \wedge \mu_{\widetilde{F}_2}(z).$$

**Proposition 2.50** $\widetilde{F}_1 * \widetilde{F}_2$ is a fuzzy interval.

However, the fuzzy arithmetic operators do not satisfy all the group properties of ordinary arithmetic. For example, in general $\widetilde{F} + (-\widetilde{F}) \neq 0$. And the use of the $\wedge$ operator in (2.49) can be (and perhaps should be) extended to $\sqcap$ (see below discussion in Sec. 5.5.2.1).

### 2.4.4 Fuzzy Sets vs. Fuzzy Measures

In the interest of being very clear about the nature of the critique of the standard understanding of the relation between possibility theory, fuzzy sets, and fuzzy measures, it is useful to consider some of the linguistic and historical issues at stake in the terminology.

There is clearly a strong linguistic similarity between terms "fuzzy set" and "fuzzy measure": presumably a fuzzy measure is a measure of some kind of fuzziness. It was certainly Sugeno's intention to closely relate the concepts of fuzzy measures and fuzzy sets. In introducing the term [277], he begins by noting that given a crisp set $F \subseteq \Omega$ and an unknown element $\omega \in \Omega$, that the "fuzziness" of the statement $\omega \in F$, denoted here as $w_\omega(F)$, should be monotonic in the cardinality of $F$:

$$F_1 \subseteq F_2 \rightarrow w_\omega(F_1) \leq w_\omega(F_2). \tag{2.51}$$

Thus $w_\omega$ should be a fuzzy measure in $F$. He then proceeds to note that $w_\omega(F)$ is just the characteristic function $\chi_F(\omega)$, but with variable $F$ and $\omega$ fixed. So, he reasons, on generalization to a fuzzy set $\widetilde{F} \subseteq \Omega$, it should follow that $\mu_{\widetilde{F}}(\omega)$ is a kind of fuzzy measure.

> Thus it can be said that the concept of grade of fuzziness in fuzzy measures theory includes as a special case the concept of grade of membership in fuzzy sets theory. [277]

However, there are a number of reasons why this choice of terms is rather poor and misleading, and historically unfortunate. First, a fuzzy measure is not the measure of the fuzziness of a fuzzy set. Sugeno's claim is rather that the fuzzy membership function is a case of a fuzzy measure. Now it is true that something like Sugeno's criteria (2.51) holds for fuzzy sets. Letting $w_\omega(\widetilde{F}) := \mu_{\widetilde{F}}(\omega)$, then from the definition of fuzzy set inclusion (2.36)

$$\widetilde{F} \subseteq \widetilde{G} \rightarrow \forall \omega \in \Omega, w_\omega(\widetilde{F}) \leq w_\omega(\widetilde{G}).$$

But this seems more an artifact of the definition of subset inclusion (2.36) than an important result in its own right.

Nor does it follow that $\mu$ is a fuzzy measure anyway. This is because $w_\omega(\widetilde{F})$ is not a set-function $w: 2^\Omega \mapsto [0,1]$ as required by the definition (2.27), but rather a *fuzzy-set* function $w: [0,1]^\Omega \mapsto [0,1]$.

Nor is a fuzzy measure the measure of the fuzziness of an element's belonging to a fuzzy set. As Sugeno admits, that is simply the membership grade itself, which is certainly not a fuzzy measure.

A number of researchers have introduced true measures of the fuzziness of a fuzzy set [155, p. 140] which *are* functions mapping fuzzy sets to the unit interval, and thus not fuzzy measures. We will continue to use the historical term "fuzzy measure" while discussing below the actual complicated relationship between fuzzy sets, fuzzy measures, and possibility distributions. But it should be kept clearly in mind that nothing of substance should necessarily follow from this historical fact.

## 2.5 Evidence Theory and Random Set Theory

The **Dempster-Shafer Theory of Evidence**, or simply **evidence theory**, of A.P. Dempster [48] and Glen Shafer [261] is somewhat less general than the theory of fuzzy measures. Nevertheless it provides a very rich domain which encompasses classical information theory and the most important classes of fuzzy measures.

Evidence theory begins with a set-based probability distribution, which we shall call an **evidence function**, and which is frequently called a **basic assignment** or **basic probability assignment**. In the following let $A, B \subseteq \Omega$.

**Definition 2.52 (Evidence Function)** [155] A function $m: 2^\Omega \mapsto [0,1]$ is an **evidence function** if $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$.

**Proposition 2.53** An evidence function $m$ induces a fuzzy set denoted $\widetilde{m} \mathrel{\widetilde{\subseteq}} 2^\Omega$ with $\mu_{\widetilde{m}}(A) := m(A)$.

Each evidence function $m$ determines two fuzzy measures from $2^\Omega$ to $[0,1]$, which we will call **evidence measures**.

**Definition 2.54 (Disjointness Relation)** Denote

$$A \perp B := A \cap B = \emptyset.$$

**Definition 2.55 (Belief Measure)** [155] Let $\mathrm{Bel}\colon 2^\Omega \mapsto [0,1]$ be a fuzzy measure where

$$\forall A \subseteq \Omega, \quad \mathrm{Bel}(A) := \Pr(\mathcal{S} \subseteq A) = \sum_{B \subseteq A} m(B).$$

**Definition 2.56 (Plausibility Measure)** [155] Let $\mathrm{Pl}\colon 2^\Omega \mapsto [0,1]$ be a fuzzy measure where

$$\forall A \subseteq \Omega, \quad \mathrm{Pl}(A) := \Pr(\mathcal{S} \not\perp A) = \sum_{B \not\perp A} m(B).$$

Bel and Pl were originally interpreted as the "lower" and "upper" probabilities, $\Pr_*$ and $\Pr^*$ respectively, defining a class of probability measures such that

$$\mathrm{Bel}(A) = \Pr_*(A) \leq \Pr(A) \leq \Pr^*(A) = \mathrm{Pl}(A). \tag{2.57}$$

**Proposition 2.58** [155] Bel and Pl are dual, in that

$$\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\overline{A}).$$

**Proposition 2.59** [155,284] Bel (and Pl as its dual) determines the evidence function according to the **Möbius inversion**.

$$m(A) = \sum_{B \subseteq A} (-1)^{|B-A|} \mathrm{Bel}(B).$$

**Corollary 2.60 (Evidence Measure Boundaries)** $\mathrm{Bel}(\Omega) = \mathrm{Pl}(\Omega) = 1$.

**Proof:**

$$\mathrm{Pl}(\Omega) = \sum_{A \not\perp \Omega} m(A) = \sum_{A \subseteq \Omega} m(A) = \mathrm{Bel}(\Omega) = 1.$$

■

The best justified combination rule for the combination of evidence functions is Dempster's rule of combination.

**Definition 2.61 (Dempster's Rule of Combination)** Given two evidence functions $m_1, m_2$ on $\Omega$, then let the **combined** evidence function be $m := m_1 \odot m_2$, where for $A_1, A_2 \subseteq \Omega$, then $\forall A \subseteq \Omega$,

$$m(A) = m_1(A) \odot m_2(A) := \frac{\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2)}{\sum_{A_1 \not\perp A_2} m_1(A_1) m_2(A_2)},$$

### 2.5.1 Random Sets

A **random set** is a random variable which takes as its values on subsets of $\Omega$. Therefore a random set $\mathcal{S}$ associates a probability, which we will denote suggestively as $m(A)$, to each $A \subseteq \Omega$. Since the $m(A)$ are probabilities and must sum to one, it is clear that under the restriction that $m(\emptyset) = 0$, then the $m$ values act as an evidence function on $\Omega$. From here on evidence functions and measures will be discussed only in the context of random sets.

Random sets were first developed in the context of stochastic geometry, dealing generally with random compact subsets of $\mathbb{R}^n$ [5,93,142,182]. It is only more recently that random sets are being integrated into the formalisms of GIT [72, 102, 181, 192].

**Definition 2.62 (Random Set)** [68] Given an evidence function $m: 2^\Omega \mapsto [0,1]$ with finite $\Omega$, a **random set** is $\mathcal{S} := \{\langle A_j, m_j \rangle : m_j > 0\}$, where $A_j \subseteq \Omega, 1 \leq j \leq N := |\mathcal{S}| \leq 2^n - 1$, and $m_j := m(A_j) = \Pr(\mathcal{S} = A_j)$.

**Definition 2.63 (Focal Set)** [155] Given a random set $\mathcal{S}$, its **focal set** is $\mathcal{F} := \{A_j : m_j > 0\}$.

**Definition 2.64 (Focal Element)** [155] Given a focal set $\mathcal{F}$, each $A_j \in \mathcal{F}$ is a **focal element**.

Note that all the $A_j$ are distinct. Since $\forall A \subseteq \Omega, A \in \mathcal{F} \leftrightarrow m(A) > 0$, therefore

$$\mathrm{Bel}(A) = \sum_{A_j \subseteq A} m(A_j), \qquad \mathrm{Pl}(A) = \sum_{A_j \not\perp A} m(A_j).$$

**Definition 2.65 (Random Set Core)** The **core** of a random set is

$$\mathbf{C}(\mathcal{S}) := \bigcap_{A_j \in \mathcal{F}} A_j.$$

**Definition 2.66 (Random Set Consistency)** [64] A random set $\mathcal{S}$ is **consistent** if $\mathbf{C}(\mathcal{S}) \neq \emptyset$.

**Definition 2.67 (Plausibility Assignment)** Given a finite random set $\mathcal{S}$ with plausibility Pl, its **plausibility assignment** (sometimes called a **falling shadow** [295], **trace** [296, 297] or **one-point coverage function** [102]), is denoted as the vector

$$\vec{\mathrm{Pl}} := \langle q_{\mathrm{Pl}}(\omega_i) \rangle = \langle \mathrm{Pl}(\{\omega_i\}) \rangle = \langle \mathrm{Pl}_i \rangle .$$

**Proposition 2.68 (Plausibility Assignment Formula)**

$$\mathrm{Pl}_i = \mathrm{Pl}(\{\omega_i\}) = \sum_{A_j \not\perp \{\omega_i\}} m_j = \sum_{A_j \ni \omega_i} m_j.$$

**Lemma 2.69** $\sum_i \mathrm{Pl}_i = \sum_j m_j |A_j|$.

**Proof:** From the plausibility assignment formula (2.68),

$$\sum_i \mathrm{Pl}_i = \sum_{i=1}^{n} \sum_{A_j \ni \omega_i} m(A_j) = \sum_{A_j \in \mathcal{F}} \sum_{\omega_i \in A_j} m(A_j) = \sum_{A_j \in \mathcal{F}} m(A_j)|A_j| = \sum_j m_j |A_j|.$$

■

**Corollary 2.70** $\sum_i \mathrm{Pl}_i \geq 1$.

**Proof:** From the Lemma (2.69), and since $|A_j| \geq 1$,

$$\sum_i \mathrm{Pl}_i = \sum_j m_j |A_j| \geq \sum_j m_j = 1.$$

■

**Theorem 2.71** If $\mathcal{S}$ is consistent, then $m(A) > 0 \to \mathrm{Pl}(A) = 1$.

**Proof:** Fix $A \subseteq \Omega$. Since $\mathcal{S}$ is consistent, $\mathbf{C}(\mathcal{S}) = \bigcap_{A_j \in \mathcal{F}} A_j \neq \emptyset$, and $\forall A_{j_1}, A_{j_2} \in \mathcal{F}, A_{j_1} \not\perp A_{j_2}$. Since $m(A) > 0$, therefore $A \in \mathcal{F}$, and so $\forall A_j \in \mathcal{F}, A \not\perp A_j$. Therefore $\mathrm{Pl}(A) = \sum_{A_j \not\perp A} m_j = \sum_{A_j \in \mathcal{F}} m_j = 1$. ■

As with the representation of a fuzzy set in the unit hypercube, a canonical, graphical representation of random sets is available. Since each random set maps evidence values, plausibilities, and beliefs on (crisp) subsets of $\Omega$, these values can be placed as labels on the vertices of the boolean hypercube of dimension $n$. An example is provided in Fig. 2.2 for a universe $\Omega = \{x, y, z\}$ and random set

$$\mathcal{S} = \{\langle\{x\}, .1\rangle, \langle\{x, y\}, .7\rangle\}, \langle\{z\}, .2\rangle\}, \qquad \mathcal{F} = \{\{x\}, \{x, y\}, \{z\}\}.$$

In the figure, nodes with value 0 are unlabeled.

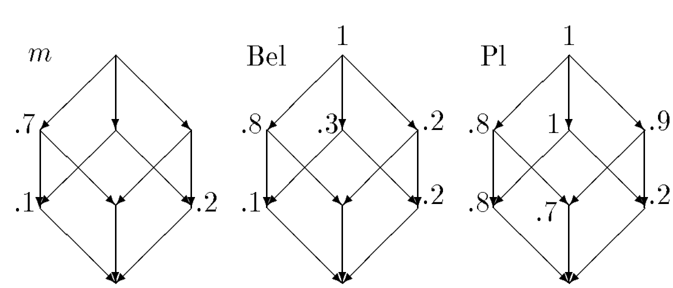


Figure 2.2: Evidence values $m$, beliefs Bel, and plausibilities Pl on $\Omega = \{x, y, z\}$.

### 2.5.2 Distributions of Random Sets

Restricting consideration to evidence measures on finite random sets, fix $\nu$ to Pl. Then the definition of distributions of fuzzy measure (2.31) is modified to random sets as follows.

**Definition 2.72 (Distribution on a Random Set)** Assume a random set $\mathcal{S}$ with plausibility Pl. The function

$$q\colon \Omega \mapsto [0,1], \qquad q := q_{\mathrm{Pl}}$$

is a **distribution** of $\mathcal{S}$ if $q_{\mathrm{Pl}}$ is a distribution of Pl. Since in general $\Omega$ and $\mathcal{S}$ are assumed to be finite, denote

$$q_i := q(\omega_i) = \mathrm{Pl}_i = \mathrm{Pl}(\{\omega_i\}), \qquad \vec{q} := \vec{\mathrm{Pl}} = \langle q_i \rangle = \langle q_1, q_2, \ldots, q_n \rangle .$$

And from (2.34), denote the fuzzy sets $\tilde{q} = \widetilde{\mathrm{Pl}} \mathrel{\tilde{\subseteq}} \Omega$.

In the sequel, the notations $q, \mathrm{Pl}$, $q_i, \mathrm{Pl}_i$ and $\vec{q}, \tilde{q}$ and $\vec{\mathrm{Pl}}, \widetilde{\mathrm{Pl}}$ may be used interchangeably as appropriate, hopefully without confusion from context.

**Proposition 2.73 (Distribution Operator)** If $q$ is a distribution of Pl on a random set $\mathcal{S}$, then from the operator condition of the distribution of a fuzzy measure (2.32),

$$\forall A \subseteq \Omega, \quad \mathrm{Pl}(A) = \bigoplus_{\omega_i \in A} q_i .$$

**Corollary 2.74 (Normalization)** If $q$ is a distribution of Pl on a random set $\mathcal{S}$, then

$$\bigoplus_{\omega_i \in \Omega} q_i = 1.$$

**Proof:** Follows immediately from the boundary conditions of evidence measures (2.60) and the definition of the distribution operator (2.73). ■

A distribution produces functional relations between the singletons and the focal elements, and between the values of the distribution and the values of the evidence function. This gives the ability to construct the focal set from the elements, and the set plausibilities from the point distribution values, and vice versa.

**Definition 2.75 (Structural Aggregation)** Given a random set $\mathcal{S}$ with a distribution $q$, then a function $g_q\colon \mathcal{F} \mapsto \Omega$ is a **structural aggregation function** if it is one to one.

A distribution not only relates the algebraic structure of $\mathcal{F}$ and $\Omega$, but also numerically aggregates the values of the $m_j$ and the $q_i$.

**Definition 2.76 (Numerical Aggregation)** Given a random set $\mathcal{S}$ with evidence function $m$, distribution $q$, and structural aggregation function $g_q$, then a function $h_{m,q}\colon [0,1] \mapsto [0,1]$ is a **numerical aggregation function** if $h_{m,q}(m(A_j)) = q(g_q(A_j))$.

Given a distribution $q$, the effect of $g_q$ is to map each focal element $A_j$ to some unique element $\omega_i$. Since $g_q$ is one to one, there can be only as many focal elements as there are elements of $\Omega$, so that

$$|\mathcal{S}| = N \leq |\Omega| = n.$$

Above, $1 \leq i \leq n$ has been used to index $\Omega$, while $1 \leq j \leq N \leq 2^n$ has been used to index $\mathcal{F} \subseteq 2^\Omega$. But when a distribution $q$ exists, the elements can be coded directly in terms of the focal elements by combining focal element and universe element notation, defining

$$\omega_j := g_q(A_j), \qquad q_j := q(\omega_j) = q(g_q(A_j)). \tag{2.77}$$

This condition will be generically called "relabeling". Thus the relabeling of (2.77) establishes a common ordering of the $A_j, m_j, \omega_i$, and $q_i$, and from numerical aggregation (2.76) it follows that

$$h_{m,q}(m_j) = q_j.$$

These relations are diagrammed in Fig. 2.3.

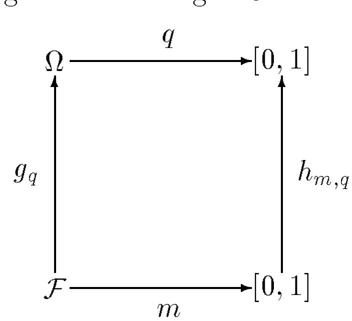


Figure 2.3: Relations among random sets and their distributions and aggregation functions.

**Definition 2.78 (Completion)** Given a random set $\mathcal{S}$ with a structural aggregation function $g_q$, then $\mathcal{S}$ and its distribution $q$ are **complete** if $N = |\mathcal{S}| = n = |\Omega|$.

**Proposition 2.79** If $q$ is complete then $g_q$ and $h_{m,q}$ are onto, and thus bijections, with

$$g_q^{-1}(\omega_j) = A_j, \qquad h_{m,q}^{-1}(q_j) = m_j.$$

So in a complete random set, the focal elements and universe elements, and the measure values and distribution values, are mutually determining, and the indices $i$ and $j$ are identical and can be used interchangeably.

### 2.5.3 Some Special Cases

Both **probability** and **possibility** arise for special cases on the structure of $\mathcal{F}$.

#### 2.5.3.1 Probability

**Definition 2.80 (Specificity)** [155] $\mathcal{S}$ is **specific** when $\forall A_j \in \mathcal{F}, |A_j| = 1$.

**Proposition 2.81** [155] If $\mathcal{S}$ is specific, then Pl = Bel and Pr := Pl = Bel is a probability measure satisfying the usual additivity conditions $\forall A, B \subseteq \Omega$

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B), \qquad A \perp B \rightarrow \Pr(A \cap B) = 0.$$

**Theorem 2.82** If $\mathcal{S}$ is specific, then $p := q_{\Pr}$ is a probability distribution with operator $\oplus = +$.

**Proof:** From the above proposition (2.81),

$$\forall A \subseteq \Omega, \quad \sum_{\omega \in A} p(\omega) = \sum_{\omega \in A} \Pr(\{\omega\}) = \Pr\left( \bigcup_{\omega \in A} \{\omega\} \right) = \Pr(A).$$

■

Thus the stochastic operation and normalization conditions are

$$\Pr(A) = \sum_{\omega_i \in A} p_i, \qquad \sum_i p_i = 1 \tag{2.83}$$

in the discrete case and

$$\Pr(A) = \int_A dp(\omega), \qquad \int_\Omega dp(\omega) = 1$$

in the continuous case.

**Theorem 2.84** If $\mathcal{S}$ is specific, then

$$g_p(A_j) := \omega_i \text{ such that } A_j = \{\omega_i\}$$

is a structural aggregation function.

**Proof:** From the definition of specificity (2.80) and the fact that the $A_j$ are all distinct, $g_p$ is clearly one to one. ■

**Theorem 2.85** If $\mathcal{S}$ is specific, then $h_{m,p}(m_j) := m_j = p_j$ is a numerical aggregation function.

**Proof:** Using the relabeling of (2.77),

$$p(g_p(A_j)) = p(\omega_j) = \Pr(\{\omega_j\}) = \sum_{A_k \subseteq \{\omega_j\}} m_k = m(\emptyset) + m(\{\omega_j\}) = m_j.$$

■

**Proposition 2.86 (Probabilistic Completion)** When $p$ is complete then $\forall \omega_i, \exists A_j = \{\omega_i\}$, and by relabeling, simply

$$g_p^{-1}(\omega_j) = A_j, \qquad h_{m,p}^{-1}(p_j) = m_j.$$

**Corollary 2.87 (Probabilistic Completion Conditions)** If $p$ is complete, then $\forall \omega_i, p_i > 0$.

**Proof:** Since $p$ is complete, then by relabeling $\forall \omega_j, \exists A_j \in \mathcal{F}$, and since $\forall A_j \in \mathcal{F}, m_j > 0$, therefore $p_j = h_{m,p}(m_j) = m_j > 0$. ■

Graphically, the specific focal elements $A_j \in \mathcal{F}$ occupy the lowest "row" of the random set, as shown in Fig. 2.4 for $\mathcal{S} = \{\langle\{x\}, .1\rangle, \langle\{y\}, .7\rangle, \langle\{z\}, .2\rangle\}$ and $\vec{p} = \langle .1, .7, .2\rangle$.

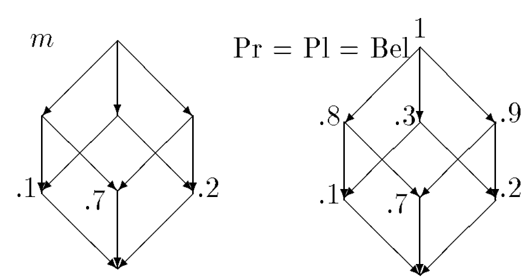


Figure 2.4: A specific random set and probability measure.

#### 2.5.3.2 Possibility

**Definition 2.88 (Consonance and Nests)** [155] $\mathcal{S}$ is **consonant** and $\mathcal{F}$ is a **nest**, when (without loss of generality for ordering, and letting $A_0 = \emptyset$) $A_{j-1} \subset A_j$.

Since the $A_j$ are all distinct, therefore $A_{j-1} \subset A_j$ and not $A_{j-1} \subseteq A_j$.

**Theorem 2.89 (Consonance/Possibilistic Equivalence)** $\mathcal{S}$ is consonant iff $\Pi := \mathrm{Pl}$ is a possibility measure.

**Proof: Case 1:** Let $A, B \subseteq \Omega$. In general, if $A \not\subseteq A_j$, then $\forall k, j \leq k \leq N, A \not\subseteq A_k$. Let $A_N$ be the maximal and $A'$ the minimal $A_j$ under the ordering $\subset$ for which $A \not\subseteq A_j$. Assume $A' \subseteq B'$. Then $(A \cup B)' = A'$, and so $\mathrm{Pl}(A \cup B) = \mathrm{Pl}(A)$. Since

$$\mathrm{Pl}(A) = \sum_{A_j = A'}^{A_j = A_N} m(A_j),$$

therefore $\mathrm{Pl}(A) \geq \mathrm{Pl}(B)$, and so $\mathrm{Pl}(A \cup B) = \mathrm{Pl}(A) \vee \mathrm{Pl}(B)$. The analogous result holds for $B' \subseteq A'$. **Case 2:** See [299, p. 64]. ∎

Of course, for a consonant random set, $\eta :=$ Bel is the dual necessity measure, as discussed in Sec. 2.2.1. Dual results for the below will not be further discussed.

**Corollary 2.90** If $\mathcal{S}$ is consonant, then $\pi := q_\Pi$ is a possibility distribution with operator $\oplus = \vee$.

**Proof:** Follows immediately from the definition of the possibility distribution (2.23) and the maximum operator for possibility theory (2.24). ∎

Thus the possibilistic operation and normalization conditions are

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i, \qquad \bigvee_i \pi_i = 1. \tag{2.91}$$

in the discrete case and

$$\Pi(A) = \sup_{\omega \in A} \pi(\omega), \qquad \sup_{\omega \in \Omega} \pi(\omega) = 1$$

in the continuous case.

In the sequel, for any given possibility distribution $\vec{\pi}$, without loss of generality let the $\pi_i$ be ordered so that

$$\pi_1 \geq \pi_2 \geq \cdots \geq \pi_n. \tag{2.92}$$

**Theorem 2.93 (Possibilistic Structural Aggregation)** If $\mathcal{S}$ is consonant, then a structural aggregation function $g_\pi : \mathcal{F} \mapsto \Omega$ exists.

**Proof:** $g_\pi$ will be specified in the course of the proof. In general, $A_{j-1} \subset A_j$, where $1 \leq j \leq N$, and $A_0 := \emptyset$ by convention, so that $A_j = A_{j-1} \cup (A_j - A_{j-1})$. Since the $A_j$ are all distinct, therefore all the $A_j - A_{j-1}$ are also distinct and non-empty. Therefore $g_\pi$ can be selected so that

$$\forall A_j \in \mathcal{F}, \quad g_\pi(A_j) \in A_j - A_{j-1}$$

arbitrarily. ∎

**Theorem 2.94 (Possibilistic Numerical Aggregation)** If $\mathcal{S}$ is consonant, then

$$h_{m,\pi}(m_j) := \sum_{k=j}^{N} m_k = \pi(\omega_j)$$

is a numerical aggregation function.

**Proof:** $\forall j$, denote $\omega_j \in A_j - A_{j-1}$ be such that from the proof of the structural aggregation theorem for possibility (2.93), $\omega_j = g_\pi(A_j)$. So $\omega_j \in A_j$, and because $\mathcal{F}$ is a nest, $\forall j \leq k \leq N, \omega \in A_k$. Therefore from the plausibility assignment formula (2.68),

$$\forall 1 \leq j \leq N, \quad \pi(g_\pi(A_j)) = \pi(\omega_j) = \sum_{A_k \ni \omega_j} m_k = \sum_{k=j}^{N} m_k = h_{m,\pi}(m_j).$$

■

**Theorem 2.95 (Possibilistic Formulae)** When $\mathcal{S}$ is complete with distribution $q = \pi$, then using the relabeling convention of (2.77)

1. $g_\pi(A_j) = A_j - A_{j-1} = \omega_j$.
2. $g_\pi^{-1}(\omega_j) = A_j = \{\omega_1, \omega_2, \ldots, \omega_j\}$.
3. $h_{m,\pi}^{-1}(\pi_j) = m_j = \pi_j - \pi_{j+1}$, where $\pi_{n+1} = 0$ by convention.

**Proof:**

1. When $\mathcal{S}$ is complete, then $|\mathcal{F}| = n$. Since all the $A_j$ are distinct, therefore $|A_j - A_{j-1}| = 1$. Since $g_\pi(A_j) \in A_j - A_{j-1}$, therefore $g_\pi(A_j) = \omega_i$ such that $A_j - A_{j-1} = \{\omega_i\}$ unambiguously. By relabeling, $\omega_j := g_\pi(A_j)$.

2. Because of the ordering convention of (2.92), the $\omega_j$ are also ordered so that

$$\pi(\omega_1) \geq \pi(\omega_2) \geq \cdots \geq \pi(\omega_n).$$

Furthermore,

$$\omega_1 = g_\pi(A_1) \in A_1 - A_0 = A_1 - \emptyset = A_1,$$

so that $A_1 = \{\omega_1\}$. Similarly,

$$\omega_2 = g_\pi(A_2) \in A_2 - A_1 = A_2 - \{\omega_1\},$$

so that $A_2 = \{\omega_1, \omega_2\}$. The result follows by induction.

3. From the completion of $\mathcal{S}$ and possibilistic numerical aggregation (2.94), $\pi_j = \sum_{k=j}^n m_k$. So
$$\pi_j - \pi_{j+1} = \sum_{k=j}^n m_k - \sum_{k=j+1}^n m_k = m_j.$$
■

In contrast with probabilistic completion (2.87), if $\pi$ is complete then all the $\pi_i$ are distinct.

**Theorem 2.96 (Possibilistic Completion)** $\pi$ is complete iff

$$1 = \pi_1 > \pi_2 > \cdots > \pi_n > 0.$$

**Proof:**

1. Assume a complete possibility distribution $\pi$.

   - $\pi_1 = 1$ from possibilistic normalization (2.26).
   - If $\exists j, \pi_j = \pi_{j+1}$ then from the possibilistic formulae (2.95), $\pi_j - \pi_{j+1} = m_j = 0$, which violates (2.62).
   - Finally, if $\exists \pi_i = 0$, then $\forall A_j, \omega_i \notin A_j$, so that $\Omega \notin \mathcal{F}$. But $\Omega \in \mathcal{F}$, because $\mathcal{S}$ is complete and consonant, and otherwise $|\mathcal{S}| < n$. Therefore $\forall \pi_i > 0$.

2. Assume a possibility distribution where $1 = \pi_1 > \pi_2 > \cdots > \pi_n > 0$. Let $i = 1$. Then
$$\pi_1 = \sum_{A_j \ni \omega_1} m_j = \sum_{A_j \in \mathcal{F}} m_j > \pi_2 = \sum_{A_j \ni \omega_2} m_j,$$
so that
$$\exists A_{j_1}, A_{j_2}, \quad \omega_1 \in A_{j_1}, A_{j_2}, \quad \omega_2 \in A_{j_1}, \quad \omega_2 \notin A_{j_2}, \qquad m_{j_2} > 0.$$
The same argument holds for general $i$, therefore $\forall 1 \leq i \leq n, \exists A_j, m_j > 0$, so that $N = n$ and $\mathcal{S}$ is complete.

■

Graphically, the nested focal elements go "up the edge" of the random set, as shown in Fig. 2.5 for $\mathcal{S} = \{\langle\{x\}, .1\rangle, \langle\{x, y\}, .7\rangle, \langle\{x, y, z\}, .2\rangle\}$ and $\vec{\pi} = \langle 1, .9, .2\rangle$.

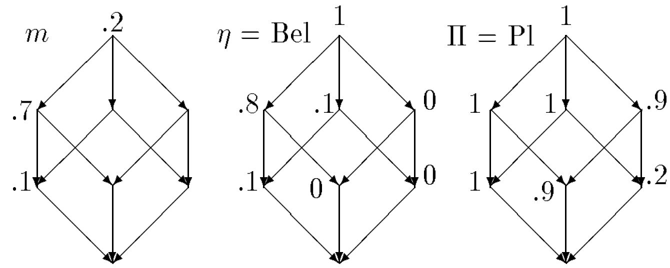


Figure 2.5: A consonant random set, possibility and necessity measures.

#### 2.5.3.3 Both

Specific and consonant random sets are almost completely distinct. There is just one set of degenerate cases that they share.

**Definition 2.97 (Certain Distributions)** The **certain distributions** $\vec{1}_i, 1 \leq i \leq n$ are those distributions for which $q_i = 1$ and $\forall k \neq i, q_k = 0$.

**Corollary 2.98** If $\mathcal{S}$ is both specific and consonant, then

$$\exists! \omega_i, \mathcal{F} = \{\{\omega_i\}\}, \qquad N = 1, \qquad \vec{p} = \vec{\pi} = \vec{1}_i.$$

**Proof:** Obvious. ■

## 2.6 Uncertainty Measures and Principles

A number of researchers have developed methods to characterize random sets and their distributions, and a variety of mathematical functions which measure the quantity and variety of evidence allocated to focal elements with respect to the algebraic structure among them. These **uncertainty measures** provide the key links between the concepts of evidence, belief, and plausibility discussed above and those of information, uncertainty, variety and constraint described in Chap. 1. The uncertainty measure of a random set or distribution quantifies the amount of freedom and variety, and lack of constraint, present in that object.

Although we define uncertainty measures on random sets, and then derive some special forms, it was the special forms that were historically almost always developed first. These special forms were almost all defined not on random sets, but rather on the distributions of fuzzy measures, and were developed to serve particular purposes in a certain methodologies or applications. The most prominent of these is stochastic entropy, which has been a crucial statistical measure since the 19th century.

### 2.6.1 Axiomatization

In general, an uncertainty measure maps a random set or probability or possibility distribution (here jointly called "objects") to $[0, \infty)$, measuring some aspect of the information or uncertainty represented by the object. When working with finite object (as here), then it is desirable to restrict the range to $[0, \log_2(n)]$.

There has recently been a great deal of work in axiomatizing uncertainty measures [154, 157, 227, 229, 231], carried out in the spirit of the axiomatization efforts for the statistical entropy measure [1]. There are many desirable properties for an uncertainty measure to satisfy. Under different possible axiomatizations, some of these properties will be axioms, and others theorems. The choice of an appropriate axiomatization is thus made on a mixture of logical, aesthetic, and arbitrary bases.

This work will not be discussed in depth here, except to (informally) mention some of the properties which are desirable for all these measures to possess [229,154].

**Symmetry:** Invariance under permutation of the values of the object.

**Expansibility:** Invariance under inclusion of additional, zero-weighted items to the object.

**Subadditivity:** The uncertainties of the projections of the object sum to no more than the uncertainty of the whole object.

**Additivity:** Equality holds when the projections are independent.

**Normalization:** There is a standard object with uncertainty value 1.

**Continuity:** Required when dealing with non-finite objects.

### 2.6.2 Uncertainty Measures on Random Sets

Two general forms of uncertainty measures on random sets have been identified.

**Definition 2.99 (Nonspecificity)** [155]

$$\mathbf{N}(\mathcal{S}) := \sum_j m_j \log_2(|A_j|),$$

The nonspecificity measures the "spread" of the evidence in $\mathcal{S}$.

**Definition 2.100 (Strife)** [154]

$$\mathbf{S}(\mathcal{S}) := -\sum_j m_j \log_2 \left[ \sum_{k=1}^{N} m_k \frac{|A_j \cap A_k|}{|A_k|} \right].$$

**Proposition 2.101** [154]

$$\mathbf{S}(\mathcal{S}) = \mathbf{N}(\mathcal{S}) - \sum_j m_j \log_2 \left[ \sum_{k=1}^N m_k |A_j \cap A_k| \right]$$

The strife measures the ambiguity in terms of the amount of discrepancy among the evidential claims $m_j$.

**Definition 2.102 (Total Uncertainty)** [154]

$$\mathbf{T}(\mathcal{S}) := \mathbf{S}(\mathcal{S}) + \mathbf{N}(\mathcal{S}).$$

**Proposition 2.103** [154]

$$\mathbf{T}(\mathcal{S}) = 2\mathbf{N}(\mathcal{S}) - \sum_j m_j \log_2 \left[ \sum_{k=1}^N m_k |A_j \cap A_k| \right].$$

Strife and nonspecificity play complementary roles in evidence theory. The former primarily measures aspects of a random set which are fundamentally probabilistic in nature, that is specific and distributed; while the latter primarily measures aspects of a random set which are fundamentally possibilistic in nature, that is nonspecific and consonant. In the general case of a random set that is neither probabilistic nor possibilistic, then the relative values of $\mathbf{N}(\mathcal{S})$ and $\mathbf{S}(\mathcal{S})$ can be consulted in order to determine the "balance" between these two kinds of information in a random set.

### 2.6.3 Uncertainty Measures on Distributions

In the cases of probability and possibility the uncertainty measures can be defined on the appropriate distributions, and they take on especially interesting forms.

#### 2.6.3.1 Probability

**Proposition 2.104** [154] If $\mathcal{S}$ is specific, then

$$\begin{aligned} \mathbf{S}(\mathcal{S}) &= \mathbf{H}(\vec{p}) := \mathbf{S}(\vec{p}) = -\sum_{i=1}^n p_i \log_2(p_i) \qquad (2.104) \\ \mathbf{N}(\mathcal{S}) &= 0 \\ \mathbf{T}(\mathcal{S}) &= \mathbf{H}(\vec{p}), \end{aligned}$$

where $\mathbf{H}$ is the stochastic entropy.

It is well-known that $\mathbf{H}$ achieves its minimum for the certain distributions, and its maximum for the equiprobable distribution

**Definition 2.106 (Maximally Uninformative Probability Distribution)** The uniform probability distribution:

$$\vec{p}^* := \langle 1/n, 1/n, \ldots, 1/n \rangle .$$

**Proposition 2.107 (Entropy Minimax)** [155]

$$\min_{\vec{p}} \mathbf{H}(\vec{p}) = \mathbf{H}(\vec{1}_i) = 0, \qquad \max_{\vec{p}} \mathbf{H}(\vec{p}) = \mathbf{H}(\vec{p}^*) = \log_2(n). \tag{2.108}$$

#### 2.6.3.2 Possibility

**Proposition 2.109** If $\mathcal{S}$ is consonant, then [154]

$$\begin{aligned}
\mathbf{N}(\mathcal{S}) &= \sum_j m_j \log_2(j) \\
\mathbf{N}(\vec{\pi}) &= \sum_{i=2}^n \pi_i \log_2\left(\frac{i}{i-1}\right) = \sum_{i=1}^n (\pi_i - \pi_{i+1}) \log_2(i) \qquad (2.110) \\
\mathbf{S}(\vec{\pi}) &= \sum_{i=2}^n (\pi_i - \pi_{i+1}) \log_2\left(\frac{i}{\sum_{k=1}^i \pi_k}\right) \\
&= \mathbf{N}(\vec{\pi}) - \sum_{i=2}^n (\pi_i - \pi_{i+1}) \log_2\left(\sum_{k=1}^i \pi_k\right) \\
\mathbf{T}(\vec{\pi}) &= \sum_{i=2}^n (\pi_i - \pi_{i+1}) \log_2\left(\frac{i^2}{\sum_{k=1}^i \pi_k}\right)
\end{aligned}$$

Whereas probability distributions have no nonspecificity, possibility distributions do have some strife. But it has been established [94] that the maximum values for $\mathbf{S}(\mathcal{S})$ for possibility measures is bounded from above, and the actual upper bound (for $|\mathcal{S}| \longrightarrow \infty$) is approximately 0.892. Hence, possibility measures are almost strife free; their strife may often be neglected, especially when $|\mathcal{S}|$ is large.

Like probability distributions, $\mathbf{N}$ achieves its minimum on certain possibility distributions, but its maximum for the distribution which is all ones.

**Definition 2.111 (Maximally Uninformative Possibility Distribution)** The unitary possibility distribution:

$$\vec{\pi}^* := \langle 1, 1, \ldots, 1 \rangle .$$

**Proposition 2.112 (Nonspecificity Minimax)** [154]

$$\min_{\vec{\pi}} \mathbf{H}(\vec{\pi}) = \mathbf{N}(\vec{1}_i) = 0, \qquad \max_{\vec{\pi}} = \mathbf{N}(\vec{\pi}) = \mathbf{N}(\vec{\pi}^*) = \log_2(n) \tag{2.113}$$

**Corollary 2.114** If $\mathcal{S}$ is consonant and $\vec{\pi} = \vec{\pi}^*$ then $\mathcal{S} = \{\langle \Omega, 1 \rangle\}$ and $\mathcal{F} = \mathbf{C}(\mathcal{S})$.

**Proof:** From the plausibility assignment formula (2.68), if $\forall i, \pi_i = 1 = \sum_{A_j \ni \omega_i} m_j = \sum_{A_j \in \mathcal{F}} m_j$, so $\forall i, \forall A_j, \omega_i \in A_j$. ∎

#### 2.6.3.3 Probability/Possibility Comparison

There have been a number of efforts over the years develop measures which compare probability and possibility distributions and measures, and conversion methods among them. In general, a **compatibility**[3] function maps a probability distribution $\vec{p}$ and a possibility distribution $\vec{\pi}$, or equivalently a probability measure Pr and a possibility measure $\Pi$, to a number in $[0,1]$, where 0 indicates complete discrepancy and 1 complete compatibility between the distributions or measures. Compatibility measures have been axiomatized by Delgado and Moral [47].

Dubois and Prade have offered a very general definition of compatibility at the measure level.

**Definition 2.115 (Dubois and Prade Compatibility (Measure))** [57] Given a probability measure Pr and a possibility measure $\Pi$, the **Dubois-Prade compatibility**, or **DP-compatibility**, of Pr and $\Pi$ is

$$\gamma_{DP}(\Pi, \Pr) = \begin{cases} 1, & \forall A \subseteq \Omega, \Pi(A) \geq \Pr(A) \\ 0, & \text{otherwise} \end{cases}.$$

If $\gamma_{DP}(\vec{p}, \vec{\pi}) = 1$ then $\vec{p}$ and $\vec{\pi}$ are **DP-compatible.**

Note that DP-compatibility is in keeping with the general ordering relation of evidence measures which interprets $\eta$ and $\Pi$ as lower and upper probabilities (2.57).

The most prominent compatibility measure was introduced by Zadeh, and will be generally sufficient in this work.

**Definition 2.116 (Zadeh Compatibility (Distribution))** [325] Given a probability distribution $\vec{p}$ and possibility distribution $\vec{\pi}$, the **Zadeh-compatibility**, or **Z-compatibility**, of $\vec{p}$ and $\vec{\pi}$ is

$$\gamma_Z(\vec{p}, \vec{\pi}) := \vec{p} \cdot \vec{\pi} = \sum_i p_i \pi_i.$$

If $\gamma_Z(\vec{p}, \vec{\pi}) = 1$ then $\vec{p}$ and $\vec{\pi}$ are **Z-compatible.**

**Theorem 2.117**

$$0 \leq \bigwedge_i p_i \leq \gamma_Z(\vec{p}, \vec{\pi}) \leq 1$$

[3] The term used in the literature is actually "consistency", so to avoid confusion with random set consistency, we will use "compatibility".

**Proof:** Since $0 \leq \pi_i \leq 1$, therefore $\forall i, 0 \leq p_i\pi_i \leq p_i$, and so

$$0 \leq \sum_i p_i\pi_i = \gamma_Z(\vec{p}, \vec{\pi}) \leq \sum_i p_i = 1.$$

But since $\pi_1 = 1$ by possibilistic normalization (2.26), therefore

$$\gamma_Z(\vec{p}, \vec{\pi}) = p_1 + \sum_{i=2}^{n} p_i\pi_i \geq p_1 \geq \bigwedge_{i=1}^{n} p_i \geq 0.$$

∎

**Lemma 2.118** If $\vec{p}$ and $\vec{\pi}$ are Z-compatible, then $\forall \omega_i$,

$$p_i > 0 \rightarrow \pi_i = 1, \qquad \pi_i < 1 \rightarrow p_i = 0.$$

**Proof:** From the proof of (2.117),

$$0 \leq \sum_i p_i\pi_i = \gamma_Z(\vec{p}, \vec{\pi}) \leq \sum_i p_i = 1.$$

Since $\forall \omega_i, p_i\pi_i \leq p_i$, therefore for equality to hold, $\forall \omega_i, p_i\pi_i = p_i$. Therefore $p_i > 0 \rightarrow \pi_i = p_i/p_i = 1$. The second result follows by *modus tolens* and the restriction $p_i, \pi_i \in [0, 1]$. ∎

**Theorem 2.119** If $\vec{p}$ and $\vec{\pi}$ are Z-compatible, then Pr determined from $\vec{p}$ and $\Pi$ determined from $\vec{\pi}$ are DP-compatible.

**Proof:** Let $A \subseteq \Omega$ be fixed. First, if $\exists \omega_i \in A, \pi_i = 1$ then

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i = 1 \geq \Pr(A).$$

On the other hand, if $\forall \omega_i \in A, \pi_i < 1$, then from (2.118) $\forall \omega_i \in A, p_i = 0$, and

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i \geq 0 = \sum_{\omega_i \in A} p_i = \Pr(A).$$

∎

### 2.6.4 Uncertainty Principles

As remarked above, uncertainty measures provide the crucial link between the formalisms of GIT and their interpretations in terms of such concepts as information, order, and organization.

As discussed in Chap. 1, the entropy measure has been advanced as an almost universally applicable explanatory concept. The **Maximum Entropy Principle**

(MEP) in particular, beginning with the maximization of thermodynamic entropy at thermodynamic equilibrium, and continuing with the Shannon-Jaynes program of informational entropy, has proved an extremely powerful formalism.

But of course, the MEP program is limited to classical information theory, and thus the special case of the strife measure in probability theory (2.1). Even in this classical context, the correlate **Minimum Entropy Principle** [34, 35] is usually neglected as a powerful tool for simplification problems.

In the face of the need to generalize the MEP to GIT, Klir has advanced three general principles for reasoning with uncertain systems [150]. These principles are formulated here in terms of a problem-solving context where for a given problem random set $\mathcal{S}$ there is a set of multiple possible solution random sets $\Gamma := \{\hat{\mathcal{S}}\}$. The task is to select an optimal solution $\mathcal{S}^* \in \Gamma$.

#### 2.6.4.1 Uncertainty Minimization

This is an arbitration principle to be used in simplification or approximation problems. It says that that solution with minimal uncertainty should be chosen so as to lose the least possible amount of information.

**Principle 2.120 (Minimum Uncertainty Principle)** [150] Given a simplification problem, let

$$\mathcal{S}^* = \min_{\hat{\mathcal{S}} \in \Gamma} \mathbf{T}(\hat{\mathcal{S}}).$$

From entropy minimization (2.108) and nonspecificity minimization (2.113), in both unconstrained probabilistic and possibility problems this principle has the same effect, selecting any of the certain distributions $\vec{1}_i$.

#### 2.6.4.2 Uncertainty Maximization

This principle is used in the context of inductive or ampliative reasoning, when it is necessary to extrapolate beyond available information. It says that that solution with maximal uncertainty should be chosen, so that it is maximally noncommittal with regard to missing information.

**Principle 2.121 (Maximum Uncertainty Principle, MUP)** [150] Given an ampliative problem, let

$$\mathcal{S}^* = \max_{\hat{\mathcal{S}} \in \Gamma} \mathbf{T}(\hat{\mathcal{S}}).$$

In a stochastic problem the MUP becomes the MEP, and in an unconstrained problem from entropy maximization (2.108), $\vec{p}^*$ will be chosen. The selection of equal probabilities as the representation of ignorance in probability theory can be traced back to Laplace's Principle of Insufficient Reason as the foundation of most methods of inductive reasoning.

In a possibilistic problem the MUP becomes the **Maximum Nonspecificity Principle**, also called the **Minimum Specificity Principle** [63, 318]. In an unconstrained problem, from nonspecificity maximization (2.113), $\vec{\pi}^*$ will be chosen.

Random sets assign essentially probabilistic evidence values $m$ to subsets $A_j \subseteq \Omega$. A number of researchers, including Dubois and Prade [57], Smets [270] and Yager [315], have suggested the following conversion formula as an application of Insufficient Reason or the MEP at the focal set level to derive a canonical probabilistic approximation to $\mathcal{S}$, replacing each subset evidence value $m(A_j)$ with the MEP uniform probability distribution over its members.

**Definition 2.122 (Maximum Entropy Probability Distribution)** [57,270,315] Given a random set $\mathcal{S}$, let the **maximum entropy probability distribution** $p^{\mathcal{S}}: \Omega \mapsto [0,1]$ be $\forall \omega \in \Omega$,

$$p^{\mathcal{S}}(\omega) := \sum_{A_j \ni \omega} \frac{m_j}{|A_j|},$$

or in vector form

$$\vec{p}^{\mathcal{S}} := \left\langle p^{\mathcal{S}}(\omega_i) \right\rangle.$$

#### 2.6.4.3 Uncertainty Invariance

Both the MUP and the Minimum Uncertainty Principle were developed in the context of a single aspect of GIT, either probability or possibility theory. But in attempting to generalize these to GIT, the need arises to convert problems from one aspect of the formalism to another. Klir introduced the **Uncertainty Invariance Principle** (UIP) [148,156] to accommodate this situation. It is used when translating a problem from one formalism to another, for example probability to possibility, and requires that the quantity of uncertainty as measured in each formalism be preserved under the transformation.

**Principle 2.123 (Uncertainty Invariance Principle, UIP)** [148] Given a translation problem where $\mathcal{S}$ is in one formalism and the $\hat{\mathcal{S}} \in \Gamma$ in another, let $\mathcal{S}^* \in \Gamma$ be such that $\mathbf{T}(\mathcal{S}^*) = \mathbf{T}(\mathcal{S})$.

## 2.7 Possibility Distributions and Random Sets

Probability and possibility clearly represent two distinct, almost isomorphic, forms for the representation of information. In the context of random sets, both probability and possibility measures are generated very naturally as two special cases of fuzzy measures with distributions. Probability is dominated by operations using $+$, while possibility is dominated by operations using $\vee$. Since probability theory has been well developed for decades, it becomes desirable to explore the properties of possibility theory, and consider both its similarities to, and still considerable differences from, probability theory.

The main value of relying on fuzzy measures with distributions is the reduction in complexity achieved when the domain of the measure is replaced by the domain of the distribution, reducing the number of states from $2^{|\Omega|}$ to $N \leq |\Omega|$. Therefore operations which take or yield distributions are far more valuable than those which take or yield measures.

### 2.7.1 Probability Distributions and Random Sets

The relation between probability distributions and random sets is simple and unproblematic: specificity of $\mathcal{S}$ is both necessary and sufficient for $\mathcal{S}$ to have an additive distribution $\vec{\mathrm{Pl}}$; and for each additively normal probability distribution $\vec{p}$ there is a unique specific random set.

**Theorem 2.124 (Probabilistic Specificity)** $\sum_i \mathrm{Pl}_i = 1$ iff $\mathcal{S}$ is specific.

**Proof:** **Case 1:** Assume relabeling convention in the probabilistic case. If $\mathcal{S}$ is specific, then $\forall A_j \neq \{\omega_j\}, \omega_j \notin A_j$, and so $\mathrm{Pl}_j = \sum_{\omega_k \in A_j} m_k = m_j$. Thus $\sum_j \mathrm{Pl}_j = \sum_j m_j = 1$. **Case 2:** If $\sum_i \mathrm{Pl}_i = 1$, then from Lemma (2.69) $\sum_j m_j |A_j| = 1$. Since $\forall A_j, |A_j| \geq 1$, therefore $\forall A_j, m_j|A_j| \geq m_j$. Since $\sum_j m_j = 1$, then it must be the case that $\forall A_j, |A_j| = 1$. ■

**Corollary 2.125** Given a complete probability distribution $\vec{p}$, then $\mathcal{S} = \{\langle \{\omega_i\}, p_i \rangle\}$ is the unique random set such that $\vec{\mathrm{Pl}} = \vec{p}$.

**Proof:** When $\vec{\mathrm{Pl}} = \vec{p}$, then $\sum_i \mathrm{Pl}_i = 1$, so from the probabilistic specificity theorem (2.124), $\mathcal{S}$ is specific. Now from the plausibility assignment formula (2.68),

$$p_i = \mathrm{Pl}(\{\omega_i\}) = \sum_{A_j \ni \omega_i} m_j.$$

In general, there will be zero or more $A_j$ such that $A_j \ni \omega_i$. If there were more than one, then one of those would have cardinality greater than one, which it cannot. If there were none, then $p_i = 0$, which it cannot from probabilistic completion (2.86). Therefore $\exists! A_j, \{\omega_i\} = A_j$, so $|\mathcal{S}| = |\vec{p}| = n = N$, and $\mathcal{S}$ is complete, and by the relabeling of (2.77), $m(A_j) = p_j$. ■

### 2.7.2 Consonance, Consistency, and Possibility Distributions

But the relationship between possibility distributions and random sets is more complicated. There is a mapping between each possibility distribution and an equivalence class of consistent random sets, although there is a unique, well-justified, consonant random set in this class.

In Sec. 2.5.3.2 it was shown that a random set $\mathcal{S}$ being consonant is both necessary and sufficient for the plausibility measure on $\mathcal{S}$ to be a possibility measure. But $\mathcal{S}$ being consonant is only a sufficient, and not a necessary, condition for $\mathcal{S}$ to have a possibility distribution.

**Theorem 2.126 (Consonance Implies Consistency)** If $\mathcal{S}$ is consonant, then $\mathcal{S}$ is consistent with core $\mathbf{C}(\mathcal{S}) = A_1$.

**Proof:** For $1 \le j \le N, A_{j-1} \subseteq A_j$. Therefore $\forall A_j, A_1 \subseteq A_j$ and so $A_1 \subseteq \bigcap_{A_j \in \mathcal{F}} A_j$. Since $A_1 \neq \emptyset$, therefore $A_1 \subseteq \mathbf{C}(\mathcal{S}) \neq \emptyset$, so that $\mathcal{S}$ is consistent. Now assume $\mathbf{C}(\mathcal{S}) \not\subseteq A_1$, then $\exists \omega_i \in \mathbf{C}(\mathcal{S}), \omega_i \notin A_1$. But since $\omega_i \in \bigcap_{A_j \in \mathcal{F}} A_j$, and $A_1 \in \mathcal{F}$, therefore $\omega_i \in A_1$, which is a contradiction. Therefore $\mathbf{C}(\mathcal{S}) \subseteq A_1$, and $\mathbf{C}(\mathcal{S}) = A_1$. ■

**Theorem 2.127 (Sufficiency of Consistency)** $\mathcal{S}$ is consistent iff $\bigvee_i \mathrm{Pl}_i = 1$.

**Proof:** **Case 1:** Assume $\mathcal{S}$ is consistent. Then $\exists \omega_i \in \mathbf{C}(\mathcal{S})$, and therefore $\forall A_j, \omega_i \in A_j$ and $\mathrm{Pl}_i = \sum_{A_j \in \mathcal{F}} m_j = 1$. **Case 2:** Assume $\bigvee_i \mathrm{Pl}_i = 1$ so that $\exists \omega_0 \in \Omega, \mathrm{Pl}(\{\omega_0\}) = \sum_{\omega_0 \in A_j} m_j = 1 = \sum_j m_j$, so that it must be that $\forall A_j, \omega_0 \in A_j$, and thus $\omega_0 \in \mathbf{C}(\mathcal{S})$, so that $\mathcal{S}$ is consistent. ■

So in possibility theory there is a disconnection between the requirements on measures and those on distributions: consonance is required for measures, and while all consonant random sets are consistent, consistency is the only *requirement* for dealing with *distributions*.

Given any of a (complete) possibility measure, possibility distribution, or consonant random set, each of the others is determined and can be constructed:

- Assume a consonant random set. Then in virtue of the equivalence of consonant random sets and possibility measures (2.89), $\Pi$ can be constructed from the definition of plausibility (2.56); and then $\pi$ is determined from the definition of the possibility distribution (2.23).

- Assume a possibility measure. Then $\pi$ is determined from the definition of the distribution (2.23); and similarly in virtue of (2.89), a consonant random set is constructed from the possibilistic formulae (2.95).

- Assume a possibility distribution. Then a a consonant random set is constructed from the possibilistic formulae (2.95); and a possibility measure from the definition of plausibility (2.91).

### 2.7.3 The Consonant Approximation

The only remaining case is when a consistent, but non-consonant, random set $\mathcal{S}$ is given. Then equally from the equivalence of consonance and possibility (2.89), Pl is not a possibility measure, so that

$$\exists A, B \subseteq \Omega, \quad \mathrm{Pl}(A \cup B) \neq \mathrm{Pl}(A) \vee \mathrm{Pl}(B).$$

But nevertheless from the sufficiency of consistency for maximum normalization (2.127), $\vec{\mathrm{Pl}}$ is maximum normalized, and can be taken as a possibility distribution $\vec{\pi} := \vec{\mathrm{Pl}}$.

**Definition 2.128 (Constructed Possibility Measure)** Assume a non-consonant, consistent random set with maximally normalized plausibility assignment $\vec{\mathrm{Pl}}$. Then let $\vec{\pi} = \vec{\mathrm{Pl}}$, and let $\Pi^*$ be the possibility measure determined by the possibilistic operator (2.91), and let $\mathcal{S}^\pi$ be the consonant random set, with focal set $\mathcal{F}^\pi$, determined by the possibilistic formulae (2.95).

In general, of course,

$$\Pi^* \neq \mathrm{Pl}, \qquad \mathcal{S}^\pi \neq \mathcal{S}.$$

**Proposition 2.129** If $\mathcal{S}$ is consistent then $\mathbf{N}(\mathcal{S}) \geq \mathbf{N}(\vec{\pi}) = \mathbf{N}(\mathcal{S}^\pi)$.

**Proof:** This is demonstrable by construction at least for $n \leq 4$, and should generalize in a complicated combinatorial proof. ∎

For a consistent, non-consonant $\mathcal{S}$, we know that $\vec{\pi} = \vec{\mathrm{Pl}}$ is a possibility distribution. But $\mathcal{S}^\pi$ is a unique, optimal, natural, and canonical consonant representation of $\vec{\pi}$. We can therefore accept *consistency* as a *sufficient* criteria to generate a possibility distribution $\vec{\pi}$ from a random set $\mathcal{S}$, and regard $\Pi^*$ as a valid representation of the measure associated with $\vec{\pi}$.

#### 2.7.3.1 Uniqueness

$\mathcal{S}^\pi$ is the unique consonant member of the equivalence class of random sets which are one-point equivalent to a given possibility distribution $\vec{\pi}$.

**Theorem 2.130** Given a possibility distribution $\vec{\pi}$, then $\mathcal{S}^\pi$ is the unique consonant random set for which $\vec{\mathrm{Pl}} = \vec{\pi}$.

**Proof:** Let $\mathcal{S}$ be a consonant random set with $\vec{\text{Pl}} = \vec{\pi}$. We need to prove that $\mathcal{S} = \mathcal{S}^\pi$. Let the $\pi_i$ be ordered as in (2.92). So from possibilistic normalization (2.26), $\pi_1 = 1$. Since from the plausibility assignment formula (2.68)

$$\pi_1 = \sum_{A_j \ni \omega_1} m_j = 1 = \sum_{A_j \in \mathcal{F}} m_j,$$

therefore $\forall A_j, \omega_1 \in A_j$. Therefore $m$ exists only on the reduced universe $\Omega' = \Omega - \{\omega_1\}$ with $|\Omega'| = n-1$. Since $\mathcal{S}$ is consonant, $\mathcal{F}$ must be a nest, and now a nest on $\Omega'$. There are $(n-1)!$ nests in $\Omega'$, one for each permutation of the $\omega_2, \omega_3, \ldots, \omega_n$. Let $\bar{\omega}_2, \bar{\omega}_3, \ldots, \bar{\omega}_n$ be the selected permutation $\bar{\Omega}'$ of $\Omega'$, and fixing $\bar{\omega}_1 := \omega_1$, then we can let $A_j := \{\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_j\}$ be the focal elements of $\mathcal{S}$. So

$$\text{Pl}_j = \sum_{A_k = A_j}^{A_k = \Omega} m(A_k) \geq \text{Pl}_{j+1} = \sum_{A_k = A_{j+1}}^{A_k = \Omega} m(A_k).$$

If $\bar{\Omega}' \neq \Omega'$, then there exists at least two $2 \leq j_1 < j_2 \leq n$ such that $\omega_{j_1} = \bar{\omega}_{j_2}, \omega_{j_2} = \bar{\omega}_{j_1}$. So $\pi_{j_1} \geq \pi_{j_2}$, but $\text{Pl}_{j_2} \geq \text{Pl}_{j_1}$. Therefore $\bar{\Omega}' = \Omega$, and so $\mathcal{F} = \mathcal{F}^\pi$. Therefore $m$ is determined from the possibilistic formulae (2.95), so that $\mathcal{S} = \mathcal{S}^\pi$. ∎

#### 2.7.3.2 Optimal Inclusion

$\mathcal{S}^\pi$ is also the consonant random set which is "closest" to $\mathcal{S}$, whether one-point equivalent or not, as developed by Dubois and Prade.

**Definition 2.131 (Random Set Inclusion)** [68] A random set $\mathcal{S}_1$ is **included** in $\mathcal{S}_2$, denoted $\mathcal{S}_1 \subseteq \mathcal{S}_2$, when $\forall A \subseteq \Omega, \text{Pl}_1(A) \leq \text{Pl}_2(A)$.

**Definition 2.132 (Optimal Inclusion)** [68] A random set $\mathcal{S}_1$ is **optimally included** in $\mathcal{S}_2$, denoted $\mathcal{S}_1 \subseteq^* \mathcal{S}_2$ when $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\mathcal{S}_1$ is the maximal such random set with respect to the partial ordering $\subseteq$.

**Proposition 2.133** [68] If $\mathcal{S}$ is consistent, then $\mathcal{S}^\pi \subseteq^* \mathcal{S}$.

#### 2.7.3.3 Natural Ordering of $\vec{\text{Pl}}$

Consider the situation where we are given only a list of uncertainty values in $[0,1]$ which are maximally normalized, that is have a maximum value of 1. While no random set (consistent, consonant, or otherwise) nor set-valued evidence measure is *apparent*, it is *justifiable* to take the list as a possibility distribution. The first, and perhaps most natural, operation that can be taken on the list is then to arrange it in order. These are the *only* conditions (maximum normalization and rank order) which are necessary for the application of the possibilistic formulae (2.95), and thus the recovery of a consonant random set and possibility measure.

#### 2.7.3.4 Canonical Random Set from a Distribution

Goodman [102] has considered the problem of constructing a random set from a given distribution, and described a method which has become somewhat canonical (e.g. for Chanas and Nowakowski [31]).

**Definition 2.134 (Canonical Random Set)** [102] Assume a general distribution $q : \Omega \mapsto [0,1]$ on a finite $\Omega$, and let $U$ be a uniform random variable on $[0,1]$. Then

$$\mathcal{S}(U) := \{\omega : q(\omega) \geq U\}$$

is a random subset of $\Omega$ with focal set $\mathcal{F}(U)$.

**Theorem 2.135** If $\vec{q}$ is maximally normalized, then $\mathcal{S}(U) = \mathcal{S}^\pi$.

**Proof:** Let $\vec{\pi} = \langle \pi_i \rangle := \vec{q}$ (finite $\Omega$), so that as before

$$1 = \pi_1 \geq \pi_2 \geq \cdots \geq \pi_n, \qquad A_j := \{\omega_1, \omega_2, \cdots, \omega_j\}, \qquad \mathcal{F}^\pi = \{A_j\}.$$

So

$$\mathcal{S}(U) = \{\omega_j : \pi_j \geq U\} = \min_{\pi_j \geq U} A_j \in \mathcal{F}^\pi.$$

Since

$$\forall A \notin \mathcal{F}, \quad m(A) = \Pr(\mathcal{S}(U) = A) = 0,$$

therefore $\mathcal{F}(U) = \mathcal{F}^\pi$. Since

$$\forall 1 \leq j \leq n, \quad U \in [\pi_j, \pi_{j+1}) \rightarrow \mathcal{S}(U) = A_j,$$

therefore

$$m(A_j) = \Pr(\mathcal{S} = A_j) = \int_{U : \mathcal{S}(U) = A_j} dU = \int_{\pi_{j+1}}^{\pi_j} dU = \pi_j - \pi_{j+1},$$

which is just the possibilistic formulae (2.95), so that $\mathcal{S}(U) = \mathcal{S}^\pi$. ∎

## 2.8 Possibilistic Normalization

In a consistent random set, all the evidential claims are in partial agreement, since they all include the core. If furthermore $\mathcal{F}$ is a nest, then $\mathbf{C}(\mathcal{S}) = A_1 \in \mathcal{F}$. Therefore a consistent random set is in some sense a "partial" nest, and it is appropriate to consider possibilistic methods on the possibility distribution $\vec{\pi} := \vec{\text{Pl}}$, the measure $\Pi^*$, and the consonant random set $\mathcal{S}^\pi$ constructed from $\vec{\pi}$.

But when $\mathcal{S}$ is not even consistent, then possibilistic concepts cannot apply in *any* capacity. Instead, possibilistic methods may only be applied to an approximation of $\mathcal{S}$ (in the measure) or $\vec{\mathrm{Pl}}$ (in the distribution). We argued in Sec. 2.7 that operations on distributions are at least as valuable as those on measures, and usually more so. In addition, the consistency requirement for a possibility distribution is weaker, and thus far easier to satisfy, than the consonance requirement for a possibility measure.

Therefore what is required are methods to approximate an inconsistent random set by a consistent one while preserving as much of the original structure of $\mathcal{S}$ as possible. Or, when approximating $\vec{\mathrm{Pl}}$, since $\bigvee_i \mathrm{Pl}_i < 1$ for an inconsistent $\mathcal{S}$, the task is to possibilistically normalize $\vec{\mathrm{Pl}}$ to produce a normal possibility distribution $\vec{\pi}$ in such a way as to preserve as much of the structure of $\vec{\mathrm{Pl}}$ as possible.

### 2.8.1 Consistent Transformations

One way to transform an inconsistent random set $\mathcal{S}$ is to move an evidential claim $\langle A, m\rangle \in \mathcal{S}$ to a new focal element. Then what is crucial is that no information be lost, and thus $\forall \omega_i \in A$ be accounted for in the transformed focal element.

**Definition 2.136 (Consistent Transformation)** A **consistent transformation** of a random set $\mathcal{S}$, denoted $\mathcal{S} \mapsto \hat{\mathcal{S}}$ with focal set $\hat{\mathcal{F}}$, evidence function $\hat{m}$, and plausibility $\hat{\mathrm{Pl}}$, moves an evidential claim $\langle A, m(A)\rangle \in \mathcal{S}$ to a focal element $\hat{A} \in \hat{\mathcal{F}}$, such that $\hat{A} \supseteq A$ in accordance with the following algorithm, where $:=$ denotes assignment:

1. $\hat{m} := m$.

2. $\hat{m}(A) := 0$.

3. $\hat{m}(\hat{A}) := \hat{m}(\hat{A}) + m(A)$.

The effect is to replace the evidence for $A$ with zero, while adding it to that of $\hat{A}$. Since $\hat{A} \supseteq A$, all the evidence of the old claim is accounted for in the new claim $\hat{A}$.

**Theorem 2.137 (Identity Consistent Transformation)** Under a consistent transformation $\mathcal{S} \mapsto \hat{\mathcal{S}}$, if $\hat{A} = A$ then $\hat{\mathcal{S}} = \mathcal{S}$.

**Proof:** Each of the steps of the algorithm has the following result:

| Step | Result |
|---:|---|
| 1 | $\hat{m}(A) := m(A), \quad \hat{m}(\hat{A}) := m(\hat{A})$ |
| 2 | $\hat{m}(A) := 0$ |
| 3 | $\hat{m}(\hat{A}) := \hat{m}(\hat{A}) + m(A)$ |

When $\hat{A} = A$, then at step 3 $\hat{m}(\hat{A}) = \hat{m}(A) = 0 + m(A) = m(A)$, with the final result that $\hat{m} = m$ and $\hat{\mathcal{S}} = \mathcal{S}$. ■

**Corollary 2.138 (Non-Identity Consistent Transformation)** Under a consistent transformation $\mathcal{S} \mapsto \hat{S}$, if $\hat{A} \supset A$, then the final result is

$$\hat{m}(A) = 0, \qquad \hat{m}(\hat{A}) = m(\hat{A}) + m(A).$$

**Proof:** Follows immediately from steps 2 and 3 of the proof of the identity consistent transformation (2.137). ■

**Theorem 2.139** $\hat{\mathcal{S}} \supseteq \mathcal{S}$.

**Proof:** Given (2.137), all we have to show is that $\hat{A} \supset A \rightarrow \hat{\mathcal{S}} \supset \mathcal{S}$. Assume that under the consistent transform

$$A \in \mathcal{F} \mapsto \hat{A} \in \hat{\mathcal{F}},$$

with $\hat{A} \supset A$, so that $m(A)$ and $m(\hat{A})$ are the only values which are changed according to (2.138). We need to show that $\forall B \subseteq \Omega, \mathrm{Pl}(B) \leq \hat{\mathrm{Pl}}(B)$.

1. If $B \not\perp A$ then $B \not\perp \hat{A}$, and since $m(A) + m(\hat{A}) = \hat{m}(A) + \hat{m}(\hat{A})$, therefore $\mathrm{Pl}(B) = \hat{\mathrm{Pl}}(B)$.

2. If $B \perp \hat{A}$, then $B \perp A$, and so $\mathrm{Pl}(B) = \hat{\mathrm{Pl}}(B)$.

3. If $B \perp A$ but $B \not\perp \hat{A}$, then $\hat{\mathrm{Pl}}(B) = \mathrm{Pl}(B) + m(A) \geq \mathrm{Pl}(B)$.

■

**Theorem 2.140** $\mathbf{N}(\hat{\mathcal{S}}) \geq \mathbf{N}(\mathcal{S})$.

**Proof:** Again from the identity consistent transformation (2.137), $\hat{A} = A \rightarrow \mathbf{N}(\hat{\mathcal{S}}) = \mathbf{N}(\mathcal{S})$. If $\hat{A} \supset A$, therefore $|\hat{A}| > |A|$, and so

$$\mathbf{N}(\hat{\mathcal{S}}) = \mathbf{N}(\mathcal{S}) - m(A)|A| + m(A)|\hat{A}| > \mathbf{N}(\mathcal{S}).$$

■

However, sometimes $\mathbf{S}(\mathcal{S}) \leq \mathbf{S}(\hat{\mathcal{S}})$, and sometimes $\mathbf{S}(\mathcal{S}) \geq \mathbf{S}(\hat{\mathcal{S}})$.

#### 2.8.1.1 Focused Consistent Transformations

So far, this definition is not very helpful, since it provides no general principle for applying the variety of consistent transformations that are available in order to derive a consistent random set.

A necessary condition for a consistent random set is a non-empty core. Since $\mathbf{C}(\mathcal{S}) \neq \emptyset \rightarrow \exists \omega_i, \forall A_j, \omega_i \in A_j$, therefore the *minimum* necessary transformation on $\mathcal{S}$ is to introduce a *minimal* core, that is a unique focus $\exists! \omega^* \in \mathbf{C}(\mathcal{S})$.

**Definition 2.141 (Focused Consistent Transformations)** A consistent transformation with focus $\omega_i := \omega^* \in \Omega$ of a random set $\mathcal{S}$, denoted $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$ with focal set $\hat{\mathcal{F}}_i$ and evidence function $\hat{m}^i$, makes $\forall A_j \in \mathcal{F}$ the consistent transformation

$$A_j \mapsto \hat{A}_j := A_j \cup \{\omega_i\} \in \hat{\mathcal{F}}_i.$$

Let $\Gamma(\mathcal{S}) := \{\hat{\mathcal{S}}_i\}$ be the family of $n$ random sets $\hat{\mathcal{S}}_i$, one for each $\omega_i \in \Omega$.

From the identity (2.137) and non-identity consistent transformation (2.138) results, for a given $A_j$, if $\omega_i \notin A_j$, then $m(A_j)$ becomes zero while the evidence for $A_j$ is added to the evidence of the "promoted" subset $\hat{A}_j = A_j \cup \{\omega_i\}$; whereas if $\omega_i \in A_j$, then there is no change.

Since $\forall \hat{A}_j \in \hat{\mathcal{F}}_i, \omega^* = \omega_i \in \hat{A}_j$, therefore all the $\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})$ are consistent with normal possibility distributions, cores $\mathbf{C}(\hat{\mathcal{S}}_i) = \{\omega_i\}$ with minimal size $|\mathbf{C}(\hat{\mathcal{S}}_i)| = 1$, and foci $\omega^* = \omega_i$.

**Theorem 2.142**

$$\forall A \subseteq \Omega, \qquad \hat{m}^i(A) = \begin{cases} m(A) + m(A - \{\omega_i\}), & \omega_i \in A \\ 0, & \omega_i \notin A \end{cases}.$$

**Proof:** Let $\omega_i = \omega^*$ and $A$ be fixed. We can consider the losses, gains, and retentions of the evidence for $A$ under the transformation $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$. The only losses will occur if $\omega_i \notin A$, in which case if $A \in \mathcal{F}$ then $m(A)$ will be lost. If $\omega_i \in A$ then $m(A)$ will be retained. Finally, $A$ will receive gains from any $A_j$ such that $A_j \cup \{\omega_i\} = A$. This is only true for $A_j = A$ (in the case of retention), or $A_j = A - \{\omega_i\}$. **Case 1:** Let $\omega_i \notin A$. Then $m(A)$ is lost, nothing is retained, and since $A - \{\omega_i\} = A$, nothing is gained. Therefore $\hat{m}^i(A) = 0$. **Case 2:** Let $\omega_i \in A$. Then there are no losses, $m(A)$ is retained, and $m(A - \{\omega_i\})$ is gained. Therefore $\hat{m}^i(A) = m(A) + m(A - \{\omega_i\})$. ■

**Theorem 2.143** $\mathcal{S} \mapsto \hat{\mathcal{S}}_i$ induces the transformation:

$$\vec{\mathrm{Pl}} = \langle \mathrm{Pl}_1, \mathrm{Pl}_2, \ldots, \mathrm{Pl}_i, \ldots, \mathrm{Pl}_n \rangle \mapsto \vec{\pi} = \langle \mathrm{Pl}_1, \mathrm{Pl}_2, \ldots, 1, \ldots, \mathrm{Pl}_n \rangle$$

**Proof:** Let $i$ be fixed so that $\omega^* = \omega_i$; let $A \in \mathcal{F}, \hat{A} \in \hat{\mathcal{F}}_i$; let $m, \mathrm{Pl}$ and $\hat{m}, \widehat{\mathrm{Pl}}$ be the evidence functions and plausibilities of $\mathcal{S}$ and $\hat{\mathcal{S}}_i$ respectively; and let $\vec{\pi}$ be the possibility distribution of $\hat{\mathcal{S}}_i$. First,

$$\pi_i = \widehat{\mathrm{Pl}}(\{\omega_i\}) = \sum_{\omega_i \in \hat{A}} m(\hat{A}) = \sum_{\hat{A} \in \hat{\mathcal{F}}} m(\hat{A}) = 1.$$

Now consider $\forall k \neq i, 1 \leq k \leq n$, and any $\emptyset \neq A_0 \subseteq \Omega$. **Case 1:** If $\omega_i \in A_0$, then $m(A_0)$ is unchanged in the transformation. **Case 2:** Assume $\omega_i \notin A_0$. If $\omega_k \in A_0$, then $\omega_k \in A_0 \cup \{\omega_i\}$; as $m(A_0)$ is added to $\mathrm{Pl}_k$, so $\hat{m}(A_0 \cup \{\omega_i\})$ is added to $\widehat{\mathrm{Pl}}_k$. Similarly, if $\omega_k \notin A_0$, then $\omega_k \notin A_0 \cup \{\omega_i\}$; as $m(A_0)$ is *not* added to $\mathrm{Pl}_k$, so $\hat{m}(A_0 \cup \{\omega_i\})$ is not added to $\widehat{\mathrm{Pl}}_k$. Therefore the transformation does not change the value of $\mathrm{Pl}(\{\omega_k\})$, and $\pi_k = \widehat{\mathrm{Pl}}_k = \mathrm{Pl}_k$. ■

#### 2.8.1.2 Choice of Focus

The task here is to transform evidence represented in the random set $\mathcal{S}$ to a consistent random set denoted $\mathcal{S}^*$, that is, assuming that $\mathcal{S}$ is not itself consistent. The method is still not complete. For a general inconsistent random set $\mathcal{S}$, there are $n$ possible focused consistent transformations $\hat{\mathcal{S}}_i$. What is still required is a method to choose the "correct" $\omega_i$ as a focus, and to elevate the plausibility of that element to 1 as a possibilistic normalization.

**Maximum Plausibility** The most obvious method is simply to select as a focus that element with the highest plausibility.

**Principle 2.144 (Maximum Plausibility)** Given a random set $\mathcal{S}$, let $\mathcal{S}^*$ be that focused, consistent transformation $\hat{\mathcal{S}}_i$ such that

$$\omega^* = \max_{\omega_i \in \Omega} \mathrm{Pl}_i.$$

This method has also been suggested by Ramer and Puflea-Ramer [232].

This is the only order-preserving focus selection.

**Corollary 2.145** Let the $\mathrm{Pl}_i$ be ordered so that

$$1 > \mathrm{Pl}_1 \geq \mathrm{Pl}_2 \geq \cdots \geq \mathrm{Pl}_n,$$

let $\vec{\mathrm{Pl}} \mapsto \vec{\pi}$ by a focused consistent transformation, and let the $\pi_i$ maintain the same indices as the $\mathrm{Pl}_i$. Then the focus $\omega^*$ selected by maximum plausibility (2.144) is the only one for which

$$\mathrm{Pl}_{i_1} \geq \mathrm{Pl}_{i_2} \rightarrow \pi_{i_1} \geq \pi_{i_2}.$$

**Proof:** By the maximum plausibility principle (2.144), $\mathrm{Pl}_1$ will be chosen as a focus. If $\mathrm{Pl}_{i_0}, i_0 \neq 1$ is changed to 1, then $\mathrm{Pl}_1 > \mathrm{Pl}_{i_0}$ but $\pi_1 < \pi_{i_0} = 1$. ■

As stressed by Dubois and Prade [59], this property is extremely important, since the ordinal relation among the $\pi_i$ is perhaps the most significant attribute of a possibility distribution (see Sec. 3.3.1).

**Minimal Information Distortion** It is also appropriate to turn to the Uncertainty Principles of Sec. 2.6.4. In particular, the choice of focus can be regarded as a transformation problem, and the UIP of Sec. 2.6.4.3 can be applied to derive $\mathcal{S}^*$ with uncertainty equal to that of $\mathcal{S}$.

**Principle 2.146** Given a random set $\mathcal{S}$, let $\mathcal{S}^*$ be that focused, consistent transformation $\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})$ such that $\mathbf{T}(\hat{\mathcal{S}}_i) = \mathbf{T}(\mathcal{S})$.

However, (2.146) cannot be used in this form. As the UIP was originally introduced [147], one side of the transformation was considered to be completely constrained, while the other was constrained only by the measure of uncertainty. For example, for a given, fixed probability distribution, the researcher would be free to select any possibility distribution with equal uncertainty.

Later results [95] show that certain transformation methods have desirable properties, but still the transformed distribution could vary with a continuous parameter $a \in (0,1)$, and it was shown that $\exists a \in (0,1)$ such that uncertainty invariance could be satisfied.

But in the present context $\Gamma(\mathcal{S})$ provides only a *finite* set of candidates from which $\mathcal{S}^*$ must be selected. If $\mathcal{S}$ is already consistent then of course $\mathcal{S}^* = \mathcal{S}$ and so $\mathbf{T}(\mathcal{S}^*) = \mathbf{T}(\mathcal{S})$. But in general it may very well be the case that $\forall \hat{\mathcal{S}}_i \in \Gamma(\mathcal{S}), \mathbf{T}(\hat{\mathcal{S}}_i) \neq \mathbf{T}(\mathcal{S})$. From (2.140) we know that $\forall i, \mathbf{N}(\mathcal{S}) \leq \mathbf{N}(\hat{\mathcal{S}}_i)$, but such a relation does not necessarily hold for $\mathbf{S}$, and therefore also not for $\mathbf{T}$. In general, there will be a tradeoff when $\mathcal{S}$ is transformed to $\hat{\mathcal{S}}_i$, with the strife of $\mathcal{S}$ being transformed into the nonspecificity of the $\hat{\mathcal{S}}_i$. But the conditions under which $\mathbf{T}(\hat{\mathcal{S}}_i)$ increases or decreases generally from $\mathbf{T}(\mathcal{S})$ have yet to be investigated.

Therefore, we must adopt a modification of this first attempt (2.146) in this finite case to express the desire to make the information contents of the original and derived random sets be as "close" as possible.

**Definition 2.147 (Distortion Function)** A function $\xi\colon \mathbb{R}^+ \mapsto \mathbb{R}^+$ is a **distortion function if**:

- $\xi(x, y) = 0$ iff $x = y$.

- $\xi(x, y)$ is monotonic increasing in $x$ and $y$.

**Principle 2.148 (Minimal Information Distortion)** Given a random set $\mathcal{S}$, let

$$\mathcal{S}^* = \min_{\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})} \xi(\mathbf{T}(\mathcal{S}), \mathbf{T}(\hat{\mathcal{S}}_i)).$$

An obvious candidate for $\xi(x, y)$ is $|x - y|$, but some other measure might sometimes be more satisfactory. Choice of a distortion function will depend on the problem at hand and the methodology of the investigator. If $\mathbf{T}(\mathcal{S}) < \mathbf{T}(\hat{\mathcal{S}}_i)$, then "extra" information is gained through the transformation that was not included in the data. On the other hand, if $\mathbf{T}(\mathcal{S}) > \mathbf{T}(\hat{\mathcal{S}}_i)$, then information in the data is lost through the transformation. In general it should be considered more dangerous to add spurious information than to excise given information, but a very great loss should not be chosen over a very small gain. One can imagine a more sophisticated loss function which would smoothly provide more weight to information loss than information gain.

There is another issue to consider here as well. To this point we have been concerned with constructing a random set $\mathcal{S}^* \in \{\hat{\mathcal{S}}_i\}$ which is a consistent approximation of an inconsistent $\mathcal{S}$, and in doing so are comparing $\mathbf{T}(\mathcal{S})$ with $\mathbf{T}(\mathcal{S}^*)$. However, the ultimate goal is to treat the plausibility assignment of $\mathcal{S}^*$ as a possibility distribution, or equivalently to derive the constructed *consonant* random set $\mathcal{S}^\pi$. If (2.129) is true, then $\mathbf{N}(\mathcal{S}^*) \geq \mathbf{N}(\mathcal{S}^\pi)$. So minimum information distortion (2.148) could instead be applied on each of the $\hat{\mathcal{S}}_i^\pi$, the consonant approximations of each of the focused consistent transformations, as follows:

**Principle 2.149 (Minimal Information Distortion, Alternate)** Given a random set $\mathcal{S}$, let

$$\mathcal{S}^* = \min_{\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})} \xi(\mathbf{T}(\mathcal{S}), \mathbf{T}(\hat{\mathcal{S}}_i^\pi)).$$

#### 2.8.1.3 An Example

Fig. 2.6 illustrates the following example. Let $\Omega = \{x, y, z\}, n = 3$, and assume the following inconsistent random set

$$\mathcal{S} = \{\langle\{x\}, .1\rangle, \langle\{x, y\}, .7\rangle\}, \langle\{z\}, .2\rangle\}, \qquad \vec{\text{Pl}} = \langle .8, .7, .2\rangle,$$

with the properties

$$\mathbf{N}(\mathcal{S}) = .7, \qquad \mathbf{S}(\mathcal{S}) = .805, \qquad \mathbf{T}(\mathcal{S}) = 1.505$$

So $\Gamma(\mathcal{S}) = \{\hat{\mathcal{S}}_x, \hat{\mathcal{S}}_y, \hat{\mathcal{S}}_z\}$. Note that $\hat{\mathcal{S}}_z$ is actually consonant, and the only $\hat{\mathcal{S}}_i$ for which $\mathbf{T}(\hat{\mathcal{S}}_i)$ increases (slightly). Maximum plausibility (2.144) selects $\mathcal{S}^* = \hat{\mathcal{S}}_x$, and

| | |
|---|---|
| $\hat{\mathcal{S}}_x = \{\langle\{x\},.1\rangle, \langle\{x,y\},.7\rangle, \langle\{x,z\},.2\rangle\}$ | $\vec{\pi}^x = \langle 1,.7,.2\rangle$ |
| $\hat{\mathcal{S}}_y = \{\langle\{x,y\},.8\rangle, \langle\{x,z\},.2\rangle\}$ | $\vec{\pi}^y = \langle .8,1,.2\rangle$ |
| $\hat{\mathcal{S}}_z = \{\langle\{z\},.2\rangle, \langle\{x,z\},.1\rangle, \langle\Omega,.7\rangle\}$ | $\vec{\pi}^z = \langle .8,.7,1\rangle$ |

Table 2.2: Example focused consistent transformations.

| | $\mathcal{S}$ | $\hat{\mathcal{S}}_x$ | $\hat{\mathcal{S}}_x^\pi$ | $\hat{\mathcal{S}}_y$ | $\hat{\mathcal{S}}_y^\pi$ | $\hat{\mathcal{S}}_z$ | $\hat{\mathcal{S}}_z^\pi$ |
|---|---|---|---|---|---|---|---|
| **N** | .700 | .900 | .817 | 1.000 | .917 | 1.209 | 1.209 |
| **S** | .805 | .317 | .259 | .269 | .224 | .363 | .363 |
| **T** | 1.505 | 1.217 | 1.076 | 1.269 | 1.141 | 1.572 | 1.572 |
| $\xi(\mathcal{S},\cdot)$ | 0.000 | .288 | .429 | .236 | .364 | .067 | .067 |

Table 2.3: Information measures of approximations.

both minimal information distortion principles ((2.148) and (2.149)), using $\xi(x,y) = |x-y|$, select $\mathcal{S}^* = \hat{\mathcal{S}}_z$.

The results are summarized in Tab. 2.2 and Tab. 2.3, where $\vec{\pi}^\omega$ is the possibility distribution of the appropriate $\hat{\mathcal{S}}_i$.

### 2.8.2 Dimensional Extension

A consistent transformation requires the modification of at least one of the values $\text{Pl}_i$, which is changed to 1 in order to possibilistically normalize $\vec{\text{Pl}}$. However, it is possible to provide a maximum normalized element in a manner which does *not* disrupt the other $\text{Pl}_i$ at all, by simply leaving them all unchanged, but instead *adding* a new element $\text{Pl}_{n+1} = 1$.

**Principle 2.150 (Dimensional Extension)** Given a possibilistically subnormal plausibility distribution $\vec{\text{Pl}}$, let $\vec{\pi}' = \langle 1\rangle + \vec{\text{Pl}}$, where $+$ in this context is vector concatenation.

The effect is to replace the universe of discourse $\Omega$ with a new universe $\Omega' = \Omega \cup \{\omega_{n+1}\}$, with a new plausibility assignment $\pi' = \vec{\text{Pl}}' = \langle \text{Pl}'_i\rangle$, where

$$\text{Pl}'_1 = 1, \qquad \text{Pl}'_i = \text{Pl}_{i-1}, \quad 2 \le i \le n+1.$$

Actually, the correct perspective is not so much that a new element $\omega_{n+1}$ is being added, as it is that a random set already defined on $\Omega'$, but for which $\exists i, \text{Pl}_i =$

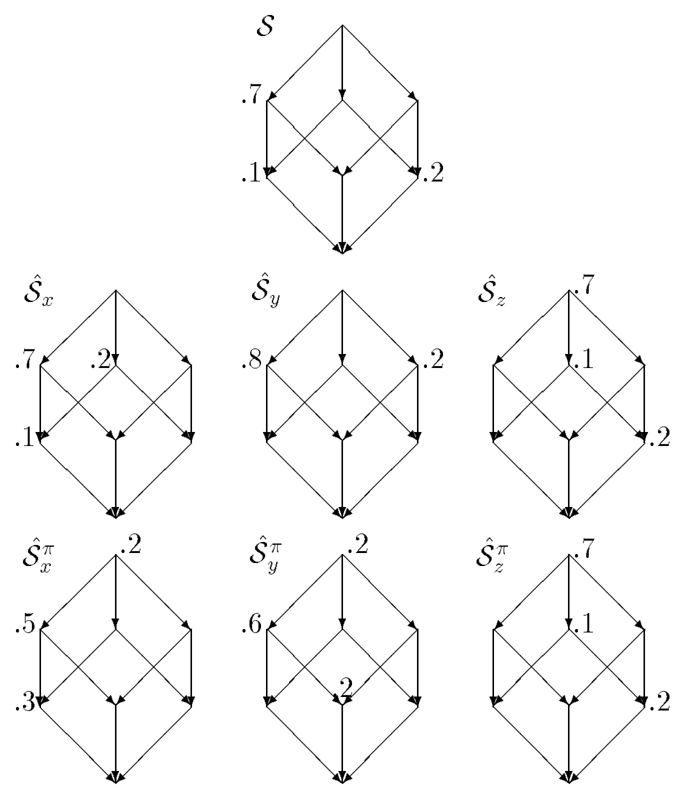


Figure 2.6: Consistent and consonant approximations of a three-dimensional inconsistent random set.

0, is consistently transformed with a focus $\omega^* = \omega_{n+1}$, effecting by (2.143) the transformation:

$$\vec{\text{Pl}} = \langle \text{Pl}_1, \text{Pl}_2, \ldots, \text{Pl}_n, 0 \rangle \mapsto \vec{\pi} = \langle \text{Pl}_1, \text{Pl}_2, \ldots, \text{Pl}_n, 1 \rangle , \tag{2.151}$$

where $\vec{\pi}$ has yet to be appropriately ordered.

**Theorem 2.152** Given an inconsistent random set $\mathcal{S}$ defined on $\Omega' = \Omega \cup \{\omega_{n+1}\}$ such that

$$\forall A_j \in \mathcal{F}, \omega_{n+1} \notin A_j, \tag{2.153}$$

then the focused consistent transformation $\mathcal{S} \mapsto \hat{\mathcal{S}}_{n+1}$ effects the transform of $\vec{\text{Pl}} \mapsto \vec{\pi}$ as in (2.151).

**Proof:** $\text{Pl}_{n+1} = \sum_{A_j \ni \omega_{n+1}} = 0$, so that $\vec{\text{Pl}}$ is as in (2.151). The result follows from (2.143), once $\vec{\pi}$ is sorted. ■

From the condition (2.153) above, $\mathcal{S}$, while technically defined on $\Omega'$, actually has weight only for $A \subseteq \Omega$, and so exists confined to the simplex $2^\Omega \subseteq 2^{\Omega'}$. Dimensional extension ((2.150) and (2.152)) projects $\mathcal{S}$ into the rest of the space involving the new element $\omega_{n+1}$.

As an example, consider the random set

$$\mathcal{S} = \{\langle \{x\}, .6 \rangle , \langle \{y\}, .4 \rangle\}$$

defined on $\Omega = \{x, y\}$ with $\vec{\text{Pl}} = \langle .6, .4 \rangle$. From dimensional extension (2.150), $\pi' = \langle 1, .6, .4 \rangle$ (once $\pi'$ is sorted) defined on $\Omega' = \{x, y, z\}$. The final random set $\hat{\mathcal{S}}_{n+1}$ is

$$\hat{\mathcal{S}}_{n+1} = \{\langle \{x, z\}, .6 \rangle , \langle \{y, z\}, .4 \rangle\},$$

as shown in Fig. 2.7. When $\mathcal{S}$ is taken to be in $\Omega'$, then $\vec{\text{Pl}} = \langle .6, .4, 0 \rangle$, as shown in the figure.

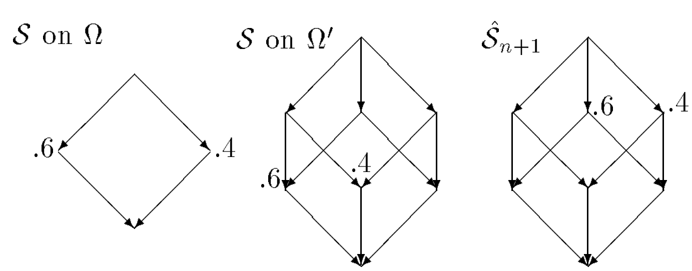


Figure 2.7: An inconsistent random set in a normal and extended universe, and its dimensional extension.

### 2.8.3 Maximum Entropy Approaches

There are other methods in the literature which have been used to select, given the values of a plausibility assignment, a random set from the equivalence class of those with that plausibility assignment (the one-point coverage problem). These methods could have some value for the current problem (selecting a consistent approximation of a given inconsistent random set). However, both of these are inappropriate for a possibilistic approach to GIT.

- When $\mathcal{S}$ is inconsistent, it might be interesting to derive a consistent approximation by considering the compatibility between $\pi$ and the maximum entropy probability distribution $p^{\mathcal{S}}$ of (2.122)

$$\mathcal{S}^* := \max_{\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})} \gamma_Z(\vec{p}^{\mathcal{S}}, \vec{\pi}^{\omega_i}).$$

  This is similar to the approach that Chanas and Heilpern [30] take to the one-point coverage problem. However, this method gives precedence not just to the probabilistic nature of $\mathcal{S}$, but its probabilistic nature based on a distortion of $\vec{\mathrm{Pl}}$ through an uncertainty principle (the MEP) which is valid only in probability theory.

- Another approach in the literature is to apply the MEP not to the $A_j$ themselves, but rather to the evidence distribution $m$ considered strictly as a probability distribution on $2^{\Omega}$. For example, the unconstrained MEP would assign uniform weight

$$\forall \emptyset \neq A \subseteq \Omega, \quad m(A) = \frac{1}{2^n - 1}.$$

  Therefore it might be interesting to consider

$$\mathcal{S}^* := \max_{\hat{\mathcal{S}}_i \in \Gamma(\mathcal{S})} \mathbf{H}(m),$$

  again similarly to Chanas and Heilpern [30]. But this approach is even less satisfactory, relying on the probabilistic MEP, but in a domain where it is completely inappropriate, and which is completely indiscriminate of any structure of the focal sets $A_j$.

## 2.9 Possibility, Probability, and Fuzzy Sets

All of the mathematical theories considered in Secs. 2.2–2.6 are related in the general context of GIT. Most of them are contained within the context of finite fuzzy measures and their specific manifestations in random sets, and thereby in evidence theory. And distributions, fuzzy measures, evidence functions, indeed, any function to $[0,1]$, define fuzzy sets, as shown for fuzzy measures and their distributions (2.34), evidence functions (2.53), and distributions on random sets (2.72).

But the history and literature of possibility theory suggests that there is a *special* relationship between possibility theory and fuzzy theory, in fact that they are almost identical. As discussed in Sec. 2.3, there is at least the obvious linguistic motivation to look for a deeper relation between fuzzy measures and fuzzy sets. And while we question Sugeno's use of the term "fuzzy" in fuzzy measures, there is clearly a relation between fuzzy sets and measures. But examination of this relation leads us to question the traditional understanding of the relation between possibility, probably, and fuzzy set theories.

### 2.9.1 Zadeh's Possibility from Fuzzy Sets

Although Shackle [260] was the first to introduce the mathematics and concepts of possibility theory, it was Zadeh [325] who first developed possibility in the context of GIT. He defined a possibility distribution *as* a fuzzy set. Beginning with a fuzzy set $\widetilde{F} \mathrel{\widetilde{\subseteq}} \Omega$, Zadeh derives a possibility distribution, here denoted $\pi_{\widetilde{F}}$, based on $\widetilde{F}$:

$$\forall \omega_i \in \Omega, \quad \pi_{\widetilde{F}}(\omega_i) := \mu_{\widetilde{F}}(\omega_i). \tag{2.154}$$

Zadeh's interpretation of possibility *strictly* in terms of fuzzy sets has *deeply* joined these ideas in the literature of GIT, to the extent that fuzzy concepts have come to dominate the work on possibility theory. Possibility distributions are interpreted strictly as "fuzzy restrictions" on variables, and the terms "possibilistic" and "fuzzy" are used synonymously. Textbooks and anthologies have been written with titles like *Fuzzy Sets and Systems, Possibility Theory, and Special Topics* [173], *Advances in Fuzzy Sets, Possibility Theory, and Applications* [294], and *Fuzzy Set and Possibility Theory: Recent Developments* [314], in which the two mathematical theories are conflated. To a very large extent, possibility theory has come to be regarded as a *branch* of fuzzy theory.

This view is unfortunate and inaccurate, and has retarded the development of both an independent possibility theory and a view of the true relation between *probability* and fuzzy sets. A deeper consideration of the relation between fuzzy sets and measures in light of the concepts of general distributions from Sec. 2.5.2 shows that neither is fuzzy theory specially related to possibility theory, nor is possibility theory the *only* form of information theory related to fuzzy sets.

### 2.9.2 Fuzzy Sets and Normalization

Since a given possibility distribution $\pi$ is a fuzzy measure distribution $q_{\mathrm{Pl}}$, therefore from (2.34), $\pi$ induces the fuzzy set $\tilde{\pi}$. However, Zadeh's definition (2.154) does *not* follow strictly from the mathematical possibility theory of Sec. 2.2 and Sec. 2.5.3.2. That is because $\pi_{\widetilde{F}}$ derived from $\widetilde{F}$ is not necessarily normal:

$$\forall \omega_i, \mu_{\widetilde{F}}(\omega_i) < 1 \rightarrow \not\exists \omega_i, \pi_{\widetilde{F}}(\omega_i) = 1.$$

Thus it is required that $\widetilde{F}$ be normal (in the fuzzy set sense of (2.41)) for $\pi_{\widetilde{F}}$ to be a possibility distribution. (2.154) in fact motivates *two* mappings, the first from fuzzy sets to possibility distributions, dependent on normalization:

$$\text{Normal}(\widetilde{F} \subseteq \Omega) \mapsto \pi_{\widetilde{F}} := \mu_{\widetilde{F}}, \tag{2.155}$$

and the second from possibility distributions to fuzzy sets, which holds in all cases:

$$\pi \mapsto \tilde{\pi} := \pi, \quad \tilde{\pi} \subseteq \Omega. \tag{2.156}$$

Instead of (2.154), what would we think if Zadeh had suggested

$$p_{\widetilde{F}}(\omega_i) := \mu_{\widetilde{F}}(\omega_i), \tag{2.157}$$

defining *probability* in terms of a fuzzy set? We would surely object, since for *almost all* $\mu$, this condition (2.157) would not yield an *actual* probability distribution. It would, however, yield an *abnormal* probability distribution, one which was either subnormal ($\sum p_i < 1$) or supernormal ($\sum p_i > 1$).

So by the same argument, even though (2.154) holds for those maximally normal $\mu$ (admittedly, for more $\mu$ than (2.157) holds), nevertheless it similarly rarely yields a maximally normal possibility distribution.

But, for those few cases when $\mu$ is in fact additively normal, then (2.157) *could* indeed hold. So if (2.154) were sufficient to define possibility, then why is it not sufficient to define probability? It cannot, since then there would be *no* restriction on what a probability distribution is: *any* string of numbers from the unit interval would be a probability distribution.

It can only be concluded that any given fuzzy set $\mu$ *could* define either a probability distribution *or* a possibility distribution, or even *both*, depending on the *properties* of $\mu$. But in no way does it follow that possibility theory is *particular* to fuzzy sets. On the contrary, it must be recognized that both probability distributions and possibility distributions are *special cases* of fuzzy sets.

The following, critical result from de Fériet [131], and translated into our notation, is useful here.

**Theorem 2.158 (de Fériet)** Assume a fuzzy set $\tilde{F} \subseteq \Omega$. Let $\Omega^+ := \mathbf{U}(\tilde{F}) = \{\omega_i\}$ be countable, and let $\langle \Omega^+, \Sigma, \nu \rangle$ be a fuzzy measure to $[0,1]$ on $\Omega^+$ such that

$$\forall \omega_i \in \Omega^+, \mu(\omega_i) = q_\nu(\omega_i) = \nu(\{\omega_i\}).$$

Then

$$\sum_i \mu(\omega_i) \geq 1 \leftrightarrow \nu = \mathrm{Pl}, \tag{2.159}$$

$$\sum_i \mu(\omega_i) \leq 1 \leftrightarrow \nu = \mathrm{Bel},$$

for some plausibility Pl and belief Bel.

**Proof:** The conditional $\leftarrow$ of (2.159) has been proved with Corollary (2.70). See [131] for the rest. ■

**Corollary 2.160** Given the conditions of de Fériet's theorem (2.158), then

$$\sum_i \mu(\omega_i) = 1 \leftrightarrow \nu = \mathrm{Pr}\,.$$

**Proof:** This has already been proved by probabilistic specificity (2.124). ■

So obviously the relation between fuzzy sets and the distributions of fuzzy measures in general, let alone possibility measures, is not as simple as Zadeh's definition (2.154) would have us believe.

As an example of the kind of equivocation resulting from interpreting a fuzzy set *necessarily* as a possibility distribution, consider a fuzzy set represented as the string of numbers $\vec{\tilde{F}} = \langle f_i \rangle = \langle .2, .7, .8, .5 \rangle$. Clearly $\bigvee f_i \neq 1 \neq \sum f_i$. Can $\vec{\tilde{F}}$ be seen as a possibility distribution? If so, then it is subnormal. But we can ask just as easily if $\vec{\tilde{F}}$ is not a *probability* distribution? If so, then it is supernormal. $\vec{\tilde{F}}$ is, of course, properly a fuzzy set. But there is no *a priori* justification to interpret it as either an abnormal possibility distribution *or* an abnormal probability distribution.

### 2.9.3 Possibilistic Normalization

It is true that maximization is a *simpler* operation than addition. But this does *not* entail that normalization in possibility theory can be overlooked any more than stochastic normalization can be in probability theory: a subnormal possibility distribution is no more a "real" possibility distribution than a sub- (or super-) normal probability distribution is a "real" probability distribution. Normalization is an *essential* feature for any general distribution of a random set, as outlined in Sec. 2.5.2.

In fact, this discussion draws into question an aspect of the axiomatization of possibility from Sec. 2.2.1. The condition $\Pi(\mathbf{0}) = 0$ from the definition of the general

| | | |
|---|---|---|
| $\tilde{F} \mathrel{\tilde{\subseteq}} \Omega$ | $\mu_{\tilde{F}}: \Omega \mapsto [0,1]$ | |
| $\tilde{m} \mathrel{\tilde{\subseteq}} 2^\Omega$ | $m: 2^\Omega \mapsto [0,1]$ | $\sum_{A \subseteq \Omega} m(A) = 1$ |
| $\tilde{\nu} \mathrel{\tilde{\subseteq}} 2^\Omega$ | $\nu: 2^\Omega \mapsto [0,1]$ | $A \subseteq B \to \nu(A) \leq \nu(B)$ |
| $\tilde{q}_\nu \mathrel{\tilde{\subseteq}} 2^\Omega$ | $q_\nu: \Omega \mapsto [0,1]$ | $\bigoplus_{\omega_i \in \Omega} q_\nu(\omega_i) = 1$ |
| $\tilde{p} \mathrel{\tilde{\subseteq}} 2^\Omega$ | $p: \Omega \mapsto [0,1]$ | $\sum_{\omega_i \in \Omega} p(\omega_i) = 1$ |
| $\tilde{\pi} \mathrel{\tilde{\subseteq}} 2^\Omega$ | $\pi: \Omega \mapsto [0,1]$ | $\bigvee_{\omega_i \in \Omega} \pi(\omega_i) = 1$ |

Table 2.4: Fuzzy sets encountered in GIT.

possibility measure (2.8) (for random sets, $\Pi(\emptyset) = 0$) is an axiom. And although it is true from the boundary conditions on evidence measures (2.60) that $\Pi(\Omega) = 1$ necessarily holds for evidence measures on random sets, in general the condition $\Pi(\mathbf{1}) = 1$ from general possibilistic normalization (2.11) ($\Pi(\Omega) = 1$ from standard possibilistic normalization (2.22)) is only a *defined* property which may or may not hold for a given measure $\Pi$. Thus the possibilistic normalization conditions of (2.91) are dependent on a normal possibility measure.

But the corresponding condition from *probabilistic* normalization (2.83) is an *axiom*, an *essential* condition. Why should it be any less so for possibility?

### 2.9.4 Fuzzy Sets and Distributions in GIT

Clearly a possibility distribution is a fuzzy set. But, in departure from Zadeh's view, possibility distributions are not the *only* fuzzy sets in the context of GIT. Consider the summary in Tab. 2.4 of the various fuzzy sets which have been proposed (from (2.33), (2.34), (2.53), and (2.72)), assuming that $\nu$ is a fuzzy measure to $[0,1]$. The conclusion is obvious and immediate: there are *many* fuzzy sets induced by the mathematical relationships in GIT, and each one has an associated condition (additivity, monotonicity, or normalization). In particular, all general distributions, including all probability and possibility distributions, are in fact fuzzy sets; all such distributions are points in the fuzzy power set $[0,1]^\Omega$. Probabilistic ($\sum \mu = 1$) and possibilistic ($\bigvee \mu = 1$) fuzzy sets are known to have special properties which make them distributions, and by which they can be mapped to fuzzy measures Pr and $\Pi$ on $2^\Omega$.

But *any* fuzzy set $\mu$ *could* be a fuzzy measure $\nu$ to $[0,1]$, an evidence function $m$, or a distribution $q$ *if the appropriate conditions hold*; in the case of distributions, if there is some fuzzy measure $\nu$ with an appropriate operator $\oplus$.

In fact, the situation is *even worse* for possibility distributions than for the other structures mentioned above. An obvious extension to de Fériet's theorem (2.158) *would be*

$$\bigvee_i \mu(\omega_i) = 1 \leftrightarrow \nu = \Pi,$$

but this is *false*.

**Corollary 2.161** Given the assumptions of de Fériet's theorem (2.158), then

$$\bigvee_i \mu(\omega_i) = 1 \not\leftrightarrow \nu = \Pi.$$

**Proof:** Clearly follows from the sufficiency of consistency for maximum normalization (2.127), assuming Pl is a plausibility on a consistent, non-consonant random set. ■

So this leads to another critique of Zadeh's definitions of possibility. There are many consistent random sets all of which have plausibility assignments equal to a given normal fuzzy set. It is only when the explicit, unjustified, and stronger (from the sufficiency of consonance for consistency (2.126)) assumption of the *consonance*, not the consistency, of the random set is included that a possibility measure is recoverable from a normal fuzzy set.

Thus (2.155) and (2.156) are insufficient, since they do not allow for the possibility of a consistent, but non-consonant, random set which generates a maximum normalized plausibility assignment equal to the fuzzy set. Instead, they allow only the recovery of the constructed possibility measure $\Pi^*$, which from Dubois and Prade's result (2.133) is only an optimal approximation to a whole class of consistent random sets.

So the relation Zadeh postulated between $\mu$ and $\pi$ does *not* entail that possibility has a privileged position with respect to fuzzy sets. In particular, a membership function can meet the requirements for many other objects in GIT, perhaps even a probability distribution.

The understanding of this deeper relation between probability and fuzziness was noticed by Kosko [164], and used by him to argue that fuzziness was not divorced from probability. In his graphical representation, probability distributions occupy points along the negative diagonal $(n-1)$–dimensional hyper-tetrahedron of the fuzzy power set. Similarly, possibility distributions occupy all hyper-faces of the hypercube which do not include the origin, as shown in Fig. 2.8 for $\Omega = \{x, y\}$ (Dubois and Prade have made a similar observation [69]). The crisp possibility distributions occupy the vertices $\langle 1, 0\rangle$ and $\langle 0, 1\rangle$.

To summarize our position, both possibility and probability measures determine unique, appropriately normalized fuzzy sets from their respective distributions. But

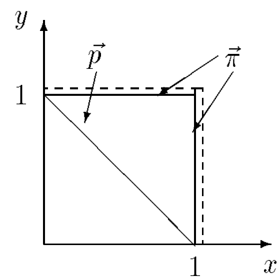


Figure 2.8: Probability and possibility distributions in the fuzzy power set.

| | Fuzzy Set = Distribution | Random Set = Fuzzy Measure |
|---|---|---|
| Probability | Additive normalized<br>On diagonal hyper-tetrahedron | Measure from evidence on singletons |
| Possibility | Maximum normalized<br>On upper hyperfaces | Measure from evidence on nest |

Table 2.5: The relations among fuzzy sets, fuzzy measures, and their distributions.

in order to derive possibility or probability measures from a fuzzy set, further conditions need apply. In the case of probability, only additive normalization is required, while in the case of possibility, maximum normalization *and* the consonance assumption is required.

Thus the overall effect is that both probability and possibility can be derived as special cases from the perspectives of either fuzzy sets or fuzzy measures, as summarized in Table 2.5. There is no necessary coupling between fuzzy sets and possibility (or probability) theory. From the perspective of possibility theory, a probability distribution is just another subnormal distribution. From the perspective of probability theory, a possibility distribution is just another supernormal distribution.

### 2.9.5 Fuzzy-Possibility Linkage

The view expressed here appears to be antagonistic to those of both the probabilistic and fuzzy communities. On the one hand, some probability theorists have exercised themselves a great deal to dismiss fuzzy theory outright [33]. While it would certainly be a retreat on their part to even admit that fuzziness was necessary, albeit only another method in the broader GIT, it would be an outright defeat for them to admit that probability *itself* can be considered as a *case* of fuzziness.

On the other hand, while GIT *practitioners* are happy to live in the mixed world of fuzziness and probability, many GIT *theoreticians* try to distance themselves from "old-fashioned" probability theory and classical information theory. Along with

Zadeh, they claim that possibility theory is the kind information theory which is "appropriate" to deal with fuzziness.

There are a number of reasons why possibility and fuzzy sets have been linked in GIT, and conversely why probability and fuzzy sets have been divorced.

#### 2.9.5.1 Historical Linkage

First and most obviously, probability theory has been in existence for many centuries, whereas both fuzzy sets and possibility theory appeared at the point of the development of GIT. Both fuzzy sets and possibility theory are departures from the classical information theory, and thus there is a desire to both group them together, and also distinguish them both from probability. Indeed, much confusion has resulting from the misinterpretation of membership grades as probability values, and a great deal of effort is taken by GIT researchers to distinguish them. It is interesting that a corresponding confusion of membership grades with *possibility* values has not troubled these researchers.

#### 2.9.5.2 Weakness of Possibilistic Normalization

Possibilistic normalization is weaker than probabilistic normalization. The measure of the number of possibility distributions on the unit hypercube of dimension $n$ is $n$, while the measure of the number of probability distributions is the measure of the hyper-tetrahedron with side length $\sqrt{2}$ and dimension $n-1$ which is less than $n$ for $n \geq 2$ (for $n = 1$ then $\langle 1 \rangle$ is both the only possibility distribution and probability distribution).

In the possibilistic normalization methods considered in Sec. 2.8, either the possibility of an element is changed to 1, leaving the others unchanged, or a unitary value is appended. Either of those methods will work for any possibilistically subnormal fuzzy set. Geometrically, dimensional extension projects a subnormal fuzzy set to unity in a direction orthogonal to all existing dimensions, while focused consistent transformations projects it to unity on one of the existing dimensions.

An example is in Fig. 2.9 for the subnormal plausibility assignment $\vec{\mathrm{Pl}} = \langle .6, .8 \rangle$ regarded as a fuzzy set in $[0,1]^{\{x,y\}}$. There are two focused consistent transformations $\vec{\pi}^x = \langle 1, .8 \rangle$ and $\vec{\pi}^y = \langle .6, 1 \rangle$. The dimensional extension is $\vec{\pi}^{n+1} = \langle .6, .8, 1 \rangle$ for $z = \omega_3$.

Possibilistic normalization is so weak that it can be easily accommodated or overlooked. But it can no more be ignored as a requirement in possibility theory than stochastic normalization can in probability theory. And while it is true that there are far more possibility distributions in the fuzzy power set than probability

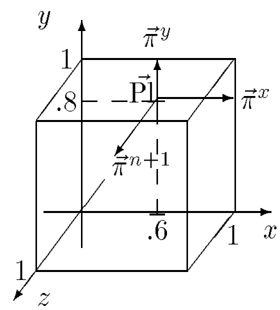


Figure 2.9: Dimensional extension and focused consistent transformation normalization.

distributions, nevertheless it is also true that there are far more fuzzy sets which are neither possibility nor probability distributions, and which fill the $n$-cube.

#### 2.9.5.3 Fuzzy Set Normalization

The criteria for a fuzzy set $\widetilde{F}$ to be "normal" (2.41), is both necessary and sufficient for $\pi_{\widetilde{F}}$ to be normal. This seems on the surface to be an historical accident in the usage of the term "normal" in each of these mathematical contexts. But the ease of maximal normalization is reinforced by the common treatment of fuzzy sets on a continuous $\Omega$, for example the ubiquitous "fuzzy numbers" of (2.44) and (2.45) [62]. Then the height of the fuzzy set $\bigvee \alpha \in \Lambda(\widetilde{F})$, is the most obvious feature of the curve, and the height being maximal is equivalent to possibilistic normalization.

In the same situation stochastic normalization results in regarding $\mu_{\widetilde{F}}$ as a probability density, so that for stochastic normalization $\int_\Omega \mu_{\widetilde{F}}(\omega)\, d\omega = 1$ would have to be satisfied. A unitary area is not an obvious, visual feature of a curve.

#### 2.9.5.4 Alpha Cuts and Focal Elements

It is suggestive that the alpha cuts $\widetilde{F}_\alpha$ of a fuzzy set $\widetilde{F}$ form a nest

$$\alpha_1 > \alpha_2 \rightarrow \widetilde{F}_{\alpha_1} \subseteq \widetilde{F}_{\alpha_2}.$$

Indeed, if $\widetilde{F}$ is normal then the $\widetilde{F}_\alpha$ are just the $A_j = \{\omega_1, \omega_2, \ldots, \omega_j\}$ constructed according to the possibilistic formulae (2.95), and $\alpha_i = \mu_i = \pi_{\widetilde{F}}(\omega_i)$. This result has been used to justify a special equivalence between a fuzzy set and a consonant random set, and thus the corresponding distribution. But there are a number of problems with this view.

First, our rejection of Zadeh's definition of possibility leads to a reasoning process which is the opposite of what is suggested above. Because of the non-determinancy of possibility measures from maximum normalized fuzzy sets, it is not justified to simply assume a nest of alpha cuts and *derive* equivalent focal elements. Rather, given a normal fuzzy set $\widetilde{F}$, and then the *assertion* that the alpha cuts are the focal elements of a random set, then it is the *possibility distribution* derived from that random set which is equivalent to $\widetilde{F}$. From the definition of alpha cuts (2.37), the ordering used in the expression $\{\omega_i : \mu_{\widetilde{F}}(\omega_i) \geq \alpha\}$ induces the ordering of the $\mu_i := \mu(\omega_i)$,

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n,$$

which is so critical to possibility theory as discussed in Sec. 2.5.3.2 and Sec. 2.8.1.2.

Furthermore, the alpha cuts of a *subnormal* fuzzy set *also* form a nest, but this does not mean that a consonant random set can be constructed from it.

**Corollary 2.162** Let $\widetilde{F}$ be subnormal with

$$1 > \mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq \mu_{n+1} = 0, \qquad A_j = \{\omega_1, \omega_2, \ldots, \omega_j\}$$

and from the possibilistic formulae (2.95), let $m(A_j) = \mu_j - \mu_{j+1}$ under relabeling. Then $\{\langle A_j, m(A_j)\rangle\}$ is not a random set.

**Proof:**

$$\sum_i m_i = \sum_i \mu_i - \mu_{i+1} = \mu_1 < 1,$$

which violates the definition of random set (2.62). ■

So the mere fact that the alpha cuts of a fuzzy set (even of a *probability distribution*) form a nest in no way lessens the normalization requirement for possibility.

Finally, while the alpha cuts of a fuzzy set can be mapped to the focal elements of a consonant random set under the conditions specified above, the focal elements of a *specific* random set can *also* be easily constructed from a fuzzy set. Indeed, because of the isomorphism between probability distributions and specific random sets from (2.125), the derivation is trivial, consisting of the singleton sets $\forall \omega_i \in \mathbf{U}(\widetilde{F}), A_i = \{\omega_i\}$. In this case, unlike in the possibilistic case, no further assumptions are required: the singleton sets compose the focal elements of a specific random set whose probability distribution is the original fuzzy set. But as with the possibilistic case, *stochastic* normalization must still hold.

**Corollary 2.163** Let $\widetilde{F}$ be such that under relabeling $\sum_j \mu_j \neq 1$, and let $A_j = \{\omega_j\}, m(A_j) = \mu_j$, as in Corollary (2.125). Then $\{\langle A_j, m(A_j)\rangle\}$ is not a random set.

**Proof:**

$$\sum_j m(A_j) = \sum_j \mu_j \neq 1,$$

which violates the definition of random set (2.62). ■

Again, neither probability nor possibility is wedded to fuzzy set theory: each is a *case* of it. A probability distribution yields a consonant class based on the ordering of the $p_i$ just as well as a possibility distribution, almost always resulting in a subnormal possibility measure. Similarly, a possibility distribution can be taken as weights on singletons as easily as a probability distribution, almost always resulting in a *supernormal* probability distribution.

#### 2.9.5.5 The Possibilistic Operator and Fuzzy Unions

Clearly the most obvious reason to conflate fuzzy theory with possibility theory is the use of the $\vee$ operator as both the possibilistic operator (2.73) and the canonical union operator (2.35) in fuzzy set theory:

$$\mu_{\widetilde{F}\cup\widetilde{G}} = \mu_{\widetilde{F}} \vee \mu_{\widetilde{G}}.$$

However, for a variety of reasons this similarity is deceptive.

In general, the fuzzy union operator (2.35) can be *any* conorm $\sqcup$, not just $\vee$. It is true that $\vee$ is canonical, and in some ways more justified than other conorms, but *formally* any will suffice, and in practice others are used.

This contrasts with both possibility theory and fuzzy measure theory. $\vee$ is the *unique* possibilistic operator; indeed, it *defines* the very domain of applicability of possibility theory, as seen in Sec. 2.2.

And while $\vee$ is indeed a conorm, in the general theory of distributions of fuzzy measure, the operator $\oplus$ need *not* be a conorm. Indeed, the other fundamental special case has $\oplus = +$, and $+$ is not a conorm.

Thus again the confusion between fuzzy and possibilistic operation is understandable, but regrettable. The point is that $\vee$ is a very weak and flexible operator. It can serve in many algebraic capacities, as both a distribution operator and a conorm. But this mere fact alone is not *prima facia* evidence to support the identification of possibility theory with fuzzy theory.

#### 2.9.5.6 Possibilistic Operations on Fuzzy Sets

It is instructive to note that many applications of fuzzy sets stress the importance of maximum normalization, of fuzzy sets having a nonempty core. For example, "fuzzy arithmetic" [139, 312] is built from fuzzy numbers (2.45), which are normal. To the extent that normalization is required, then indeed these fuzzy numbers yield to a possibilistic interpretation. But then since the essence of possibility theory is the requirement of maximum normalization, by parsimony it is more appropriate to regard such "fuzzy mathematics" as a branch of *possibility* theory proper, rather than as an application of *fuzzy* theory.

## 2.10 Probability and Possibility

In establishing a GIT, it is important to understand the parts that compose it, their relations to each other, and their historical development. In particular, we are interested in the status of possibility theory as a part of GIT, and in comparison with probability theory both as a part of GIT and as the essential core of classical information theory.

The questions that arise include: Are possibility and probability values necessarily related? Is possibility more general than probability? Do they express a form of duality, symmetry, opposition, or independence? Many of the factors which affect these questions are decidedly semantic in nature. Thus they are not inherent in the formalism, and will be considered more fully in Chap. 3. Some of the formal aspects of this problem will be considered here, with the conclusion that while probability and possibility are complementary and symmetric, nevertheless possibility expresses a weaker form of uncertainty than probability.

### 2.10.1 Probabilistic and Possibilistic Independence

One view is provided by Klir, and strongly supported by the development of Sec. 2.5.

> [Probability and possibility theory] are not comparable. Neither of them is more general than the other. In fact, either theory represents a particular set of bodies of evidence for each given $\Omega$, and these two sets do not overlap except for the special body of evidence that describe a situation with full certainty (full information): $m(\{\omega\}) = 1$ for some particular $\omega \in \Omega$. [153, p. 16]

Klir's final observation is reflected directly by the graphical representation of general distributions in Fig. 2.8. The probability and possibility distributions can be seen as "orthogonal", overlapping only at the atoms of the lattice. These distributions correspond to all the certain distributions $\vec{1}_i$, on which the two normalization conditions coincide, so that $\sum q_i = \bigvee q_i = 1$, as shown in Sec. 2.5.3.3.

A further representation of this property can be seen in Fig. 2.4 and Fig. 2.5. In a stochastic random set the focal elements occupy the lowest row of the power set, the atoms of the boolean lattice. In a possibilistic random set the focal elements occupy one of the $n!$ possible chains from the lattice 1 to lattice 0. All of these chains can be seen as "orthogonal" to the atoms.

Thus on a *purely* formal view, probability and possibility are virtually independent. They refer to distinct kinds of uncertainty, each with its own internal logic and rules of operation. Within GIT they form a complementary pair. This view is *strongly* supported by the development of Sec. 2.5.3 especially, where possibilistic and stochastic properties are developed in parallel, with almost completely symmetrical results.

### 2.10.2 Possibilistic Weakness

But on consideration of other factors, it can be seen that while probability and possibility are formally complementary, nevertheless possibility expresses a less constraining form of uncertainty than probability (see also a discussion of this issue by Dubois and Prade [75]). Thus while possibility is not more *general* than probability, it is *weaker*. This weakness can actually be of great *value*, because it allows reasoning about uncertainty without the excess constraints which probability theory brings.

This weakness in comparison with probability is manifested in a number of ways:

**Strictness of Probability:** Probability is a very special case of an evidence measure. In general, in evidence theory the focal elements, the only subsets on which positive evidence is presented, are generic subsets of $\Omega$ with varying cardinalities. Under these conditions the evidence measures establish the interval $[\mathrm{Bel}, \mathrm{Pl}]$ with $\mathrm{Bel} \leq \mathrm{Pl}$. But also Bel and Pl are dual under (2.58).

Under the strictest conditions, the cardinalities of the focal elements collapse to unity, and this interval $[\mathrm{Bel}, \mathrm{Pl}]$ collapses to a point, so that $\mathrm{Bel} = \mathrm{Pl}$, and $\mathrm{Pr} = \mathrm{Bel} = \mathrm{Pl}$, and the duality is lost. Thus probability is the *strictest* form of evidence measure, and every *proper* belief and plausibility is a weaker fuzzy measure than a probability. Since possibility is a proper plausibility, possibility also is weaker than probability.

Lastly, under a "meta-probabilistic" interpretation of a random set as a true probability distribution on subsets of a universe,[4] the occurrence of a probability distribution as the distribution of the random set appears as the special case where there is a collapse between the two levels of the evidence function as a distribution on sets and the measure distribution as a distribution on singletons.

[4] This idea has actually been little explored, perhaps only by Fung and Chong [90].

**Ease of Normalization:** Another reflection of the weakness of possibility is the ease with which a fuzzy set can be possibilistically normalized, as described in Sec. 2.9.5.2. Given a generic fuzzy set $\widetilde{F}$, an $\omega_i \in \Omega$ is selected, and $\mu_i$ is raised to 1; or simply a 1 is appended.

This is contrasted with the difficulty of stochastic normalization by such minimal modifications of $\widetilde{F}$. In general, for $n \geq 2$, if $\sum \mu_i > 2$, then a minimum of $\lceil \sum \mu_i \rceil - 1$ values will require modification. Of course, unlike possibility, this method of stochastic normalization is not well justified. But stochastic normalization methods that are more appropriate for probability distributions are far more complex, and in general require modification of *all* $\mu_i$, completely distorting the information in $\widetilde{F}$. Thus stochastic normalization is much harder to satisfy, making the ease of possibilistic normalization all the easier to dispense with by comparison.

**Number of Focal Sets:** For a fixed universe of discourse, there are vastly more consonant (let alone consistent) than specific random sets. Given $|\Omega| = n$, then there is exactly one complete specific focal set $\mathcal{F} = \{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}\}$. However, as noted in Sec. 2.10.1, there are $n!$ complete nests.

**Possibility as Maximally Weak:** (2.28) establishes upper and lower bounds on the fuzzy measure of the intersection and union of two subsets of the universe. This bound is achieved only in the cases of necessity and possibility respectively, and therefore possibility (and necessity in the dual) are the weakest possible fuzzy measures.

# Chapter 3

# Semantics of Possibility Theory

The existence of an invariant over a set of
phenomena implies a constraint, for its
existence implies that the full range of variety
does not occur. As every law of nature implies
the existence of an invariant, it follows that
every law of nature is a constraint.
— *W. Ross Ashby*

The mathematical systems we construct are enticing in their elegance and beauty. It is easy for mathematicians to revel in their complexities, and lose sight of their dependencies and limits. It is easy to commit referrential fallacies, confusing the meaning for the token, the territory for the map. It is easy to focus completely on our symbol strings, and lose sight of the underlying processes of measurement and interpretation. This is something we must avoid by clearly developing a well-justified semantics for possibility theory.

The history of modern science can be seen as a ongoing, interrelated development of mathematical formalism and scientific theory. Advances in mathematical systems are matched by their applications in science, in turn furthering mathematical development. This reciprocal relationship can be seen as a relation between a mathematical **syntax**, or the formal properties of mathematical systems; and a scientific **semantics**, or the interpretations and meanings of those formalisms. Together, syntax, semantics, and pragmatics (the study of the *use* of formal systems) compose the field of **semiotics** [15, 78, 79, 188, 259], which can be described as the general science of symbol systems.

This dissertation can be seen as a work in the semiotics of possibility theory. Chap. 2 concerned the mathematical syntax of possibility theory, and the pragmatics of possibility will be considered in Chaps. 6 and 7.

The semantics of possibility theory will occupy Chaps. 3 through 5. First, in this chapter some of the requirements for a coherent, objective semantics for possibility theory that can be used to ground applications of possibility theory to natural science will be outlined. We will begin by considering the general nature of **semantic relations** in semiotic systems, in particular the importance of **measurement** procedures in the application of formal systems. After some discussion of the logical and probabilistic criteria for possibilistic semantics, the variety of concepts and representations which inhere in possibility theory will be examined. Finally we will critique traditional semantics of possibility, including **subjective estimation**, **converted frequency** methods, possibility as **statistical likelihood**, and **objective measurement** methods for fuzzy sets.

## 3.1 Possibility Theory in the Semiotics of Modeling Relations

Our basis for the semantics and interpretation of possibility theory will rest on ideas from semiotics about sign-functions and codings, and the isomorphic ideas from Systems Science about models. These concepts will be central to the development of an objective possibilistic semantics in this chapter, to possibilistic measurement methods in Chap. 4, and to possibilistic models in Chap. 5.

### 3.1.1 Symbol and Meaning

One of the key points from semiotics concerns the general nature of the relation between a **symbol** and its **meaning**. First, following the early semioticians de Sassure [254] and Peirce [206], a symbol is understood as a relation between a **signifier**, which is the physical marker or **token**; and a **signified**, or **referent**, which is a general phenomena, otherwise called the **meaning** of the token. Together, the signifier and the signified form what de Sassure calls a **sign-function**. In another sense it can be said that the token **represents** the signified in virtue of the **coding** of the sign-function.

This formulation is deceptively simple. In fact, in the definition of the sign function one of the primary manifestations of the mind-body problem can be seen, where a material token is related to a possibly immaterial "meaning". The **problem of reference**, which has occupied philosophers of language for centuries (but not us here specifically), results: what is the nature of the sign-function, of the coding relation, of reference and representation? How is it established, manifested, maintained? What does "meaning" itself mean?

### 3.1.2 The Semantics of Modeling Relations

Very similar ideas arise in cybernetics and systems science in the general theory of **models**. The basics of modeling theory are well established.

**Definition 3.1 (General Model)** [149,241,290] Assume sets $W = \{w\}$ and $M = \{m\}$; an **object system** $S_1 = \langle r, W \rangle, r: W \mapsto W$; a **modeling system** $S_2 = \langle f, M \rangle, f: M \mapsto M$; and a **coding function** $o: W \mapsto M$. Then $O := \langle S_1, S_2, o \rangle$ is a **model** if $r, f$, and $o$ form a homomorphism, so that $\forall w \in W, o(r(w)) = f(o(w))$.

A general model is shown in Fig. 3.1, where the diagram must commute.

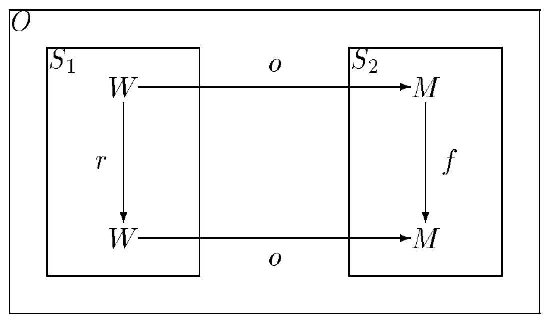


Figure 3.1: A general modeling relation.

We can say that within the model $O$ the modeling system $S_2$ models the object system $S_1$. Semiotically, $o$ is a sign-function between the $w \in W$ and the $m \in M$, so that given $m = o(w)$, then $m$ represents the referent $w$ in virtue of the coding $o$. Thus the coding $o$ serves a semantic function within the model, encoding the meaning of the $w$ in terms of the $m$, while the functions $r$ and $f$ serve syntactic functions, transferring the $w$ and the $m$ through some dynamical processes into the future.

#### 3.1.2.1 Formal Semantics

The nature of $M$ and $W$ have yet to be specified. In the history of semantics, it is common for both $S_1$ and $S_2$ to be formal systems, so that both the $m$ and $w$ are tokens. This results in models which effectively *translate* expressions between these formal systems. Many examples are possible, including the rewrite rules of formal languages and the theory of "denotational semantics". Under these conditions, the symbols $m$ are translations of the $w$, and effectively serve as *meta-symbols*.

This is unsatisfactory here because the effect is simply to create yet another formal system: the formalism $S_2$ is simply subsumed into another model at a higher level. None of the essential features of the use of mathematics for scientific modeling, where the referents are real systems, is captured.

It should be noted that exactly such a formal approach is what is developed in the so-called theory of measurement [167]. This is concerned with models where the modeled system $S_1$ is a general formal system, and within the modeling system $S_2$, $M$ is a set of subsets of $\mathbb{R}^n$. The theory is then concerned with the formal properties of the various possible homomorphisms between $S_1$ and $S_2$. Yager [311] has applied this formal measurement theory to possibility theory.

#### 3.1.2.2 Symbol Grounding in Natural Semantics

Natural science is concerned with the case where $S_1 = \langle W, r \rangle$ is a system of causal, ontological entailments, an aspect of the natural world. In this context the semantic coding function of $o$ is understood as the *processes* of **measurement** [25, 200, 202].

In a measurement procedure some aspect of the natural world enters into interaction with a **measuring device**, perhaps as simple as a counting mechanism, or as complex as an elaborate scientific instrument, resulting in a symbol $m$ of the formalism. Then **natural law** can be described as the establishment of a modeling relation between the causal system $S_1$ and the inferential system $S_2$ which brings them into congruence through the coding $o$.

Here Fig. 3.1 takes the form of Fig. 3.2. Both $S_1$ (the world) and $S_2$ (the model) are now parameterized in time, where $t$ and $t'$ indicate prior and subsequent time. The measurement $o_t = o(w_t)$ at time $t$ is used to **instantiate** the model. Then the output of the model $m_{t'} = f(m_t) = f(o_t)$ is **corroborated** against the measurement $o_{t'} = o(w_{t'})$. If these are equal, then it is a good model.

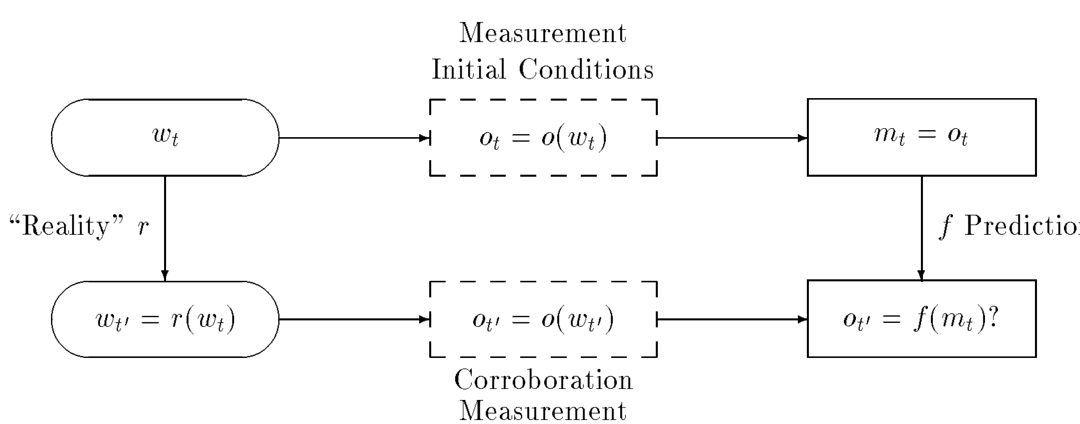


Figure 3.2: Models in natural science.

So in natural science the problem of reference becomes the problem of establishing a good measurement function $o$ from the natural system $S_1$ to the formalism $S_2$. In the research programs of artificial intelligence and artificial life, the problem of reference has been described as the "symbol grounding problem" [117, 199]. These disciplines look to the ways in which encoding and decoding occurs in real systems. It has been suggested, by Rosen [242] among others, that all living systems fall into this class, being biosemiotic systems which through their metabolic and reproductive processes are constantly involved with interpretation. More particularly, the scientific process itself can be seen as such a modeling relation, where $S_2$ is simply the body of scientific theory, and $S_1$ the world in general.

### 3.1.3 Freedom and Constraint in Semiotic Systems

One of the crucial, and seemingly paradoxical, aspects of sign-functions, and thus of models, is the relation between token and meaning seen in terms of freedom and constraint. On the one hand, this relation is **arbitrary**. Meanings of expressions are not, and cannot be, inherent in the formalism. Rather they are *constrained* by the formalism to take on only those meanings which are consistent with the formalism. In the most general semiotic systems, when $S_2$ has no extra structure, then we arrive at the hallmark property of symbols: that interpreting agents are completely free to take any token to have any meaning. But on the other hand, once a given symbol is *used in a model*, its meaning must remain *fixed*. Thus within one system of representation there is the mixed, necessary presence of both ultimate freedom and ultimate constraint.

This can be seen in the modeling formalism in Fig. 3.1. While the entailments $r$ and $f$ within $S_1$ and $S_2$ are given and fixed, the coding relation $o$ between $W$ and $M$ itself is not entailed by them. Multiple codings are possible, whether $O$ is a good model or not. From the perspective of either $S_1$ or $S_2$, the other system is completely free. If in fact the coding $o$ composes a good model, so that $O$ commutes, then the system $O$ as a whole becomes fixed, but the *parts* of $O$ are still free *with respect to each other.*

This bears out the *semantic* nature of $o$: there is no *necessary* relation between the entailments within $W$ and $M$ and that of the coding. Thus semantic codings, in particular measurement procedures in scientific models, are a central concern for any formal method, and will be in particular for a well-grounded possibility theory.

### 3.1.4 Meta-State Representations of Uncertainty

Our purpose here is to substitute possibility theory for $S_2$. So the formalism $S_2$ is deterministic, but we want it to represent nondeterminism and uncertainty. Are our efforts doomed?

This question comes up routinely, particularly in introductory discussions of fuzzy theory. What, people ask, could this apparently oxymoronic term "fuzzy logic" possibly mean? Are the rules of fuzzy logic fuzzy? What value is there in fuzzy rules? (It is instructive to note that people do not (at least any more) ask the same about probability: are the rules of probability random?)

Of course, all representations of uncertainty are themselves certain. Instead, a theory of uncertainty stands as a "formalism translation" model as discussed in Sec. 3.1.2.1. They all encode uncertainty in terms of *meta-states* (for example, probability distributions, possibility distributions, or fuzzy sets), and then propose deterministic rules that act on them at this meta-level. They do not "capture" actual uncertainty, but rather succeed in "pushing certainty back one level". Just as the laws of probability are not random, so the laws of fuzziness are not vague, and the laws of possibility are not imprecise.

This has been remarked on quite well by Goodman and Nguyen.

> ... in order to prevent a possibly infinite regress of nested multi-truth evaluations, the meta-level of description must be classical. Perhaps someday (the authors are unaware of any work in this direction), serious textbooks will be published which might typically include: "Pr(Theorem$A$) = 0.3", or "Pr(Pr(Theorem$A$) $>$ $x$) $=$ $q(x)$, for most $x \in B$", with all "proofs" developed through multi-valued or multi-multi-valued logic. However, until that time arrives, we must be content with this apparent paradoxical situation. [104, p. xvii]

In the modeling language, the system $S_2$ is not a simple function, but is itself a model of uncertainty in terms of these meta-states. Thus in all models of uncertainty, including possibility theory, $O$ is not a simple model, but rather a **meta-model**.

This type of state-space recursion is quite common, for example in the treatment of nondeterministic automata in machine theory, or the probabilistic representation of states in quantum theory. Both automata theory and quantum physics are quite deterministic at this higher level.

### 3.1.5 Models with Uncertainty

Finally we arrive at the presentation of models with uncertainty, and possibilistic models in particular. Figs. 3.3 and 3.4 show stochastic and possibilistic models respectively.

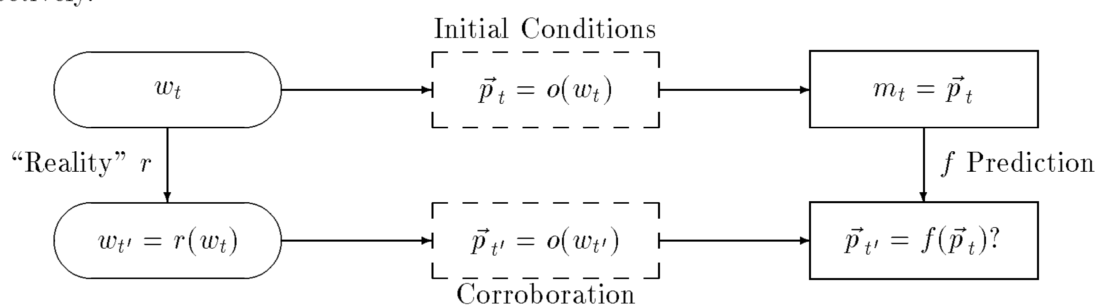


Figure 3.3: A stochastic model.

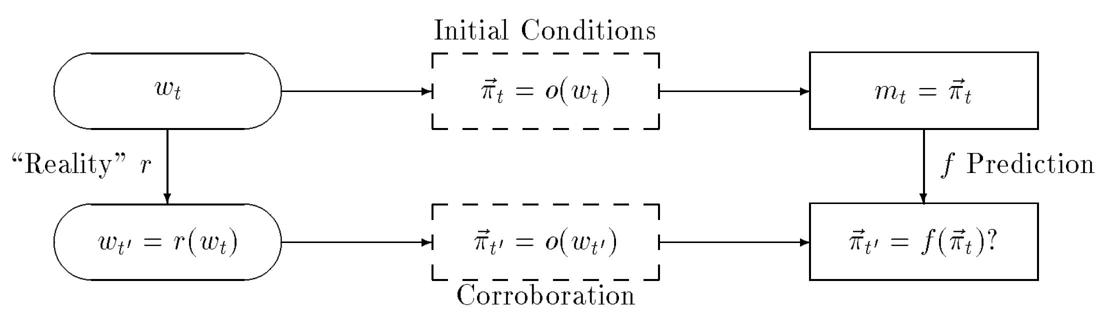


Figure 3.4: A possibilistic model.

Here $M = \{\vec{p}\}$, a set of probability distributions, or $M = \{\vec{\pi}\}$, a set of possibility distributions, respectively. Thus each meta-state of the model represents the probability or possibility of the world existing in each of the states $w \in W$.

The measurement procedures of initialization and corroboration, as well as prediction methods, must also be modified to the cases with uncertainty. Stochastic measurement has been well established for many decades. Standard methods include curve-fitting to frequency histograms, maximum likelihood estimation, etc. Stochastic prediciton methods are also well known, including Markov processes and Bayesian networks.

The corresponding concepts in possibility theory are significantly underdeveloped. For measurement, in the vast majority of approaches in the literature, $o$ is simply a person who provides a subjective judgment. The general character of this problem will occupy much of this chapter, and the technical details for possibilistic measurement will be presented in Chap. 4. Possibilistic prediction methods are equally scarce in the literature. Possibilistic processes, including possibilistic Markov processes, will be defined in Chap. 5.

## 3.2 Criteria for a Possibilistic Semantics

As discussed in Sec. 3.1.3, potential interpretations of a formal theory are not determined, but are only constrained, by the specifics of the formalization. Within those constraints, however, we are free to construct whatever interpretations we choose which are consistent with the mathematics. In deriving a semantics of possibility, we must be cognizant of the history and alternative development of these concepts, and adhere to them wherever possible.

The concept of possibility and the closely related concepts of necessity, impossibility, and contingency have a long history in philosophy. There is discussion from Aristotle on about the meaning of possibility and the kinds and varieties of possibility and impossibility.

In this section we will consider criteria which restrict the semantics of possibility, including general logical criteria, crisp modal logics, and the relation between possibilistic and probabilistic mathematics.

### 3.2.1 Graduated Possibility

The fundamental question for the semantics of possibility has classically been "What could it mean to be *possible*?" In exploring the semantics of *mathematical* possibility theory, as outlined in Chap. 2, the question becomes "What could it mean to be *somewhat* possible?". This is the central issue for mathematical possibility theory, and for which it is unique: the interpretation not of "crisp" possibility, but possibility which admits to *degrees*, specifically the representation of graduated possibility in terms of possibility distributions.

The philosophy of possibility is almost exclusively dedicated to crisp possibility. Very few have considered the specific question of graduated possibility. Giles offers a brief musing.

> To assert '$A$ is impossible' is at least as strong as, and perhaps for all practical purposes equivalent to, asserting '$\neg A$'. The assertion '$A$ is possible' is more ambiguous: does the speaker mean '$A$ is *just* possible' or '$A$ is entirely possible' or something in between? The first statement implies that $A$ is unlikely; the second seems to entail no commitment at all — whether $A$ is, or is not, found to hold the speaker cannot be

> contradicted. This ambiguity suggests that we should ... admit 'degrees of possibility', say ranging from zero (impossible) to one (entirely possible). [98]

But the first, and still most extensive, serious consideration of the meaning of degrees of possibility was by Shackle [260]. He equates a degree of possibility with a degree of "surprise."

> This state or act of mind [that rejects that a given thing will happen], expressed in other words, is a judgment that the thing in question is *impossible*. The occurrence of something hitherto judged impossible would cause a man a degree of surprise which is the greatest he is capable of feeling. If this be so, we have, corresponding to *perfect possibility*, a zero degree of surprise; corresponding to *impossibility*, an *absolute maximum* degree of surprise. Can there be *degrees of possibility*? ... There are degrees of surprise. If surprise corresponds to possibility, then we can say that there are degrees of possibility. The greatest surprise is caused by the occurrence of the impossible. If a lesser degree of surprise occurs, it must surely be because the occurrence was judged not quite impossible ... A very slight surprise indicates something which the individual had little difficulty in imagining to come true. Surprise provides us with a means of knowing how strongly we doubted the possibility of a given happening or a given outcome of some act of our own. [260, p. 68, italics original]

Shackle then qualifies his position by defining possibility as *potential* surprise, since uncertainty statements, such as probability statements, must concern some expectation of future events.

If we identify the "degree of potential surprise" with $1 - \Pi$, Shackle develops an axiomatic system which is essentially equivalent to that of mathematical possibility theory, including the dual nature of belief in the context of possibility (2.10), the independence of the possibility values of a statement and its complement (2.18), the maximum operator (2.24), and joint and conditional possibility (see Chap. 5).

Shackle also laid out many of the semantic consequences of mathematical possibility theory. For example (all of these will be discussed later in this chapter), he emphasizes the superior representation of complete ignorance, the independence of possibility from the cardinality of the universe, and the corresponding dependence of probability on the establishment of a fixed frame.

Possibility, and particularly the concept of "equi-possibility", has been used historically in the philosophy of probability beginning with Leibniz and Laplace. They were not concerned with developing a full theory of possibility *per se*, but rather with finding a mechanism to motivate the fundamental postulates of probability. But, as argued by Hacking [112, Chapter 14] and Reichenbach [235], even this effort is seriously flawed.

### 3.2.2 Physical Possibility

Hacking [111, 112], in his modern survey of the philosophical issues surrounding possibility, traces the fundamental distinction between *de re* and *de dicto* possibility (literally the possibility "of things" and "of statements" respectively) back to the Scholastics. This distinction reflects the similar distinction at the heart of probability theory, where objective, aleatory, ontological, "physical" probabilities are determined from frequencies; and subjective, epistemic, "mental" probabilities are determined from (typically Bayesian) estimation. In fact, various views of this distinction have resulted in the various schools of philosophy of probability and a vast literature (see [83] for an introduction).

This distinction is also a central concern for this work, since GIT is dominated by subjective methods for determining e.g. fuzzy membership values, evidence values, and possibilities (a subject which will occupy Sec. 3.4). Our interest is to help construct the methodological and conceptual basis for *de re*, graduated possibility.

Hacking reflects this distinction by contrasting the phrases "it is possible that $X$ happened" and "it is possible for $X$ to happen".

> The first possibility is relative to our state of knowledge and has long been called epistemic. The second possibility says it is physically possible for [$X$, nothing prevents it]. 'Possible that' tends to be epistemic (unless preceded by the adverb 'logically'), while 'possible for' goes with actual abilities independent of our knowledge of them. [112, p. 123]

While graduated (as opposed to crisp) possibility has not been especially interesting to philosophers, *physical* (as opposed to *subjective*) possibility has been. Clearly the classical founders of probability, in their use of the concept of "equipossibility" to ground probability theory, resorted (at least sometimes) to possibility in the physical sense. In other words, it was argued that it was objectively *observable* which were the (equally) possible cases in order to count frequencies or make estimations about their probability.

A classical example from Laplace concerns a biased coin where $\Pr(H) = a \neq \Pr(T) = b, a \neq 0 \neq b$, but it is not known whether $a > b$ or $b > a$. So since

$$a > b \rightarrow a^2 > ab > b^2, \qquad a < b \rightarrow a^2 < ab < b^2,$$

therefore no matter whether $TT$ or $HH$ is expected, neither $TH$ or $HT$ is expected. Laplace says

> One regards two events as equally probable when one can see no reason that would make one more probable than the other, because, even though there is an unequal possibility between them, we know not which way, and this uncertainty makes us look on each as if it were as probable as the other [112, quotation on p. 132].

Here Laplace is referring to graduated, *de re* possibility. It is graduated, since while the possibilities are unequal, clearly they must also both be *somewhat* possibility, and thus non-zero. And it is *de re*, since the unfairness of the coin is a property of the coin *itself*, rather than our *knowledge* of the coin.

There is an *additional* uncertainty about the unfairness of the coin, whether $T$ or $H$ should be favored. This may in fact be a *de dicto* uncertainty, but is an uncertainty which is whether to favor $TT$ or $HH$, and not in the fact that both $TH$ and $HT$ should *not* be most favored.

But of course Laplace was not interested in pursuing a theory of possibility further. The task of this work is to delineate what kind of meaning there is in possibility values which are both physical, *de re*, and graduated, admitting to degrees.

### 3.2.3 Modal Possibility

In modern philosophy possibility has been formally represented in the modal logics of the early 20th century [124, 169]. Modal logic is a relatively simple extension to propositional logic. Modal operators on a proposition $p$ are available, $M(p)$ for the statement "$p$ is possible" and $L(p)$ for "$p$ is necessary". These are related by a duality similar to that of mathematical possibility and necessity (2.10)

$$M(p) = \neg L(\neg p), \qquad L(p) = \neg M(\neg p). \tag{3.2}$$

From there a variety of additional axioms are available, and a variety of classes of modal logics follow.

Modal logic since Aristotle has been a *crisp* theory: propositions are explicitly possible, necessary, impossible or contingent. "Quantified" modal operators have been introduced to extend predicate logic, allowing expressions such as $\forall x, L(p(x))$. But no *multi-valued* or *graduated* modal logic has been introduced to allow expressions of the form $L_{.5}(p)$ or $L(p) = .5$, meaning "$p$ is half-necessary".

There have been some attempts to unite modal and fuzzy measure possibility [217, 239, 244]. Dubois and Prade [70] suggest an obvious translation of the basic terms

$$L(p) := \eta(p) = 1, \qquad M(p) := \Pi(p) = 1,$$

and a more interesting definition of necessary implication

$$L(p \to q) := (\eta(p \to q) = 1) \to (p \to q).$$

The prevailing semantic basis of modal logic is the "many-worlds" interpretation. This can be summarized as saying that propositions can be evaluated in multiple contexts ("worlds"), that necessity indicates truth in all contexts, and possibility truth in at least one context. An obvious extension to quantify modal operators would be to introduce a relative measure of the number of true worlds to false worlds, but none of them take this approach.

Although this entire subject deserves further consideration, it can be said that modal logic generally remains a highly mathematical form of philosophy, with little or no application in science, and, in particular, no measuring or interpretation procedures. It provides a robust mathematical theory of crisp possibility, but little more.

### 3.2.4 Natural Language Possibility

Aside from the formalism of possibility theory, the criteria for possibilistic semantics are also constrained by purely logical and methodological considerations. In particular, we are not free to ignore the results of the above modal possibilistic concepts, and ordinary language "common sense".

For example, we look at a six-sided die and say that there are six possible outcomes of a toss, so that $\Omega = \{1, 2, 3, 4, 5, 6\}$. This is clearly a crisp sense of possibility: the expression is not modified or quantified, the various options are just *possible*, that is with possibility 1. We may then question whether the die is fair or not, and consider the distribution of occurrences of the various faces. But this is then embarking on a probabilistic analysis: there is never any question that each face is *completely* possible, no matter how unevenly *likely* they may be.

Another factor in the common language usage of crisp possibility is the relation of possibility to occurrence. One definition of "possibility" offered by Webster is "being something that may or may not occur [301]". Näther observes: "The popular meaning of possibility [is]: events which take place at least one time are possible (but not necessarily probable)" [193]. The conclusion is obvious: something that actually *happens* must be (or have been) *possible*. This property will be crucial in the following development: if an event $A \subseteq \Omega$ is observed to occur, then $\Pi(A) = 1$.

But just because an event has *not* occurred does not mean that it is *not* possible, perhaps even *completely* possible. On the contrary, some event may be possible, but simply has not occurred *yet*. The die may be hidden to us before it is rolled, and each roll may produce a new face not previously seen. Just because a five has not yet appeared does not mean that a five is *not* possible, only that it *may or may not* be possible. But once all six faces have been observed at least once, then they must be given unitary possibility. Therefore $0 < \Pi(A) \leq 1$ means that $A$ is possible, that is, not prohibited, but also not necessarily *seen*.

Finally, consider the use of the common English phrase that "$X$ is not a probability, but it is a possibility," or "$X$ is possible, but not probable." How can this be interpreted? First, whereas we have begun with a non-problematic sense of crisp possibility, and are trying to derive a graduated sense, when interpreting this phrase we begin with a *graduated* sense of *probability*, and are searching for a *crisp* one: what does it mean to be just "probable" or "improbable"? On first consideration, it would seem to mean $\Pr > .5$ or $\Pr < .5$ respectively. Then the possibilistic clause of the phrase could be interpreted as $\Pi = 1$. But how is it justified that crisp probability simply means "likely", so that $\Pr > .5$? If $\Pr(A) = .49$, is $A$ not "probable"? Or does $\Pr(A)$ have to reach .75, .90, or .99 before it could be called probable? Can any crisp cutoff be justified? If not, then why cannot "$A$ is probable" simply mean $\Pr(A) > 0$? Or finally, perhaps the phrase could simply mean that while $A$ is a possibility, so that $\Pi(A) = 1$ or $\Pi(A) \gg 0$, that $\Pr(A)$ is somehow *indeterminate*?

### 3.2.5 Probabilistic Criteria

As discussed in Chap. 2, possibility exists as part of a greater GIT which includes other specific formalisms. Possibility stands in closest relation to probability, both being examples of fuzzy measures with distributions. While we intend to develop possibility theory as a kind of information theory distinct from probability, there is no desire to jettison the wealth of history which probability theory provides us. Therefore in moving away from crisp possibility to consideration of the semantics of graded possibility, it is appropriate to take the semantics of probability into account.

#### 3.2.5.1 Gaines-Kohout Compatibility

We begin with a remark of Kosko, who reminds us that "After the fact 'randomness' looks like fiction [164]." Here he is referring to the well-known "paradox" of probability statements: if after flipping the coin a heads is observed, what sense does it make to say that the probability of heads is one half? Probability statements are necessarily *predictive*, statements of what is to be, not what is. With increasing time, there is less to predict, and their value is dissipated. Some suggest that after the occurrence of an event $A$, it should realistically be said that $\Pr(A) = 1$. Surely then it must also be that $\Pi(A) = 1$, in keeping with the observation from Sec. 3.2.4 that occurrence implies complete possibility.

In their seminal articles on possibility theory [91, 92], Gaines and Kohout go further in considering the status of possibility and probability statements over time. They observe that a "likely" event $A$, here interpreted as $\Pr(A) > 0$, actually has the property of *eventuality:* with increasing time the *aggregate* probability of $A$ occurring during that time approaches arbitrarily close to one. Their discussion deserves quotation at length.

> In our studies of system stability and control we have been very concerned to embody in our formulation the distinction between possible events that *may not occur* and possible events that are *guaranteed to occur* sooner or later. The former events correspond to problems that may arise and have to be avoided. They relate to regions of states which are reachable in terms of stability analysis but not reachable in terms of control. The second type of possible event, however, is responsive to feedback control since if the situation is continually re-created in which it may occur then it eventually *will* occur.
>
> Note that probability theory does not provide an explicatum of the first type of possible event. If for the purposes of analysing an uncertain system we assign an uncertain event a non-zero probability then we imply that not only may it occur but also, in a sequence of occurrences each of which may be that event, it eventually *will* occur with a probability arbitrarily near one. The notional assignment of a definite probability to an event also fails to provide an adequate event because it has the stronger implication that the relative frequency of such events in a sequence will tend to converge to the given probability with increasing length of sequence.
>
> Either or both of these connotations which probability has over possibility may be too strong in practical situations where the concepts of probability theory are being used to express the effects of uncertain behavior. For example, we are often faced with situations where an event $E$ *may* occur, but there is no guarantee that $E$ actually *will* occur, no matter how long we wait. If we ascribe some arbitrary probability to $E$ then we certainly express that it is a possible event. However we are in a position to derive totally unjustified results based on the certainty of

> some eventual occurrence of $E$, or meaningless numeric results based on the actual "probability" of occurrence of $E$. [92]

Kosko also notices this property: "Unlike fuzziness, probability dissipates with increasing information [164]". That is, at any *particular* time the event $A$ *could* happen, and as time passes, in the limit $A$ becomes *certain* to occur.

Gaines and Kohout attempted to merge probabilistic and modal possibilistic concepts by distinguishing among so-called "eventual" and merely "possible" events. While their concept of possibility is still crisp, nevertheless we can draw from their observations.

In Sec. 3.2.4 it was said that the occurrence of an event $A$ must mean that $\Pi(A) = 1$. So by Gaines and Kohout's reasoning, this idea should be extended to say that if an event $A$ *must* occur *sometime*, that is, is *eventual*, then similarly it should be that $\Pi(A) = 1$.

The following definition will be important in considering this relation of possibility and probability.

**Definition 3.3 (GK-compatibility)** A function $f$ is **Gaines-Kohout-compatible**, or just **GK-compatible**, with a function $g$ on a set $X = \{x\}$ if $f, g \colon X \mapsto [0,1]$ and

$$\forall x \in X, \quad f(x) > 0 \leftrightarrow g(x) = 1.$$

**Corollary 3.4** If $f$ is GK-compatible with $g$ on $X$ then

$$\forall x \in X, \quad f(x) = 0 \leftrightarrow g(x) < 1.$$

**Proof:** Follows trivially from the restrictions $f(x), g(x) \in [0,1]$. ■

The following principle then relates possibility and probability measures.

**Principle 3.5 (Probability-Possibility Compatibility (PPC))** Given a probability measure Pr and possibility measure $\Pi$ then Pr is GK-compatible with $\Pi$ on $2^{\Omega}$.

Simply stated, $\forall A \subseteq \Omega$,

$$\Pr(A) > 0 \leftrightarrow \Pi(A) = 1, \qquad \Pr(A) = 0 \leftrightarrow \Pi(A) < 1. \tag{3.6}$$

This states that something having non-zero probability is, following Gaines and Kohout, *likely*, and therefore *eventual*, and therefore equivalent to its being *completely possible*. Conversely, a properly possible event ($\Pi(A) < 1$) must be of probability measure zero, and probability zero may or may not indicate proper possibility.

If a probability and possibility measure are GK-compatible, then so are their (discrete) distributions.

**Theorem 3.7 (Distribution Compatibility)** If Pr is GK-compatible with $\Pi$ on $2^\Omega$ then $p$ is GK-compatible with $\pi$ on $\Omega$.

**Proof:** $p: \Omega \mapsto [0,1]$ and $\pi: \Omega \mapsto [0,1]$, so the first condition of (3.3) is satisfied. **Case 1:** Assume $\forall \omega \in \Omega, p(\omega) > 0 \not\rightarrow \pi(\omega) = 1$. Then $\exists \omega, p(\omega) = a > 0, \pi(\omega) = b < 1$. Then $\Pr(\{\omega\}) = a, \Pi(\{\omega\}) = b$, which violates the GK-compatibility of Pr with $\Pi$ on $2^\Omega$. Therefore $\forall \omega \in \Omega, p(\omega) > 0 \rightarrow \pi(\omega) = 1$. **Case 2:** Assume $\forall \omega \in \Omega, \pi(\omega) = 1 \not\rightarrow p(\omega) > 0$. Then $\exists \omega, \pi(\omega) = 1, p(\omega) = 0$. Then $\Pi(\{\omega\}) = 1, \Pr(\{\omega\}) = 0$, which violates the GK-compatibility of Pr with $\Pi$ on $2^\Omega$. Therefore $\forall \omega \in \Omega, \pi(\omega) = 1 \rightarrow p(\omega) > 0$. ∎

**Corollary 3.8** If Pr is GK-compatible with $\Pi$ on $2^\Omega$ then $\eta(A) > 0 \rightarrow \Pr(A) > 0$, and $\eta(A) = 0 \rightarrow \Pr(A) = 0$.

**Proof:** If $\eta(A) > 0$, then by (2.19), $\Pi(A) = 1$, and so by PPC, $\Pr(A) > 0$. If $\eta(A) = 0$, then by (2.19), $\Pi(A) < 1$, and so by PPC, $\Pr(A) = 0$. ∎

Thus an event may have zero probability, and yet still have some degree of necessity, or a positive probability and no degree of necessity.

#### 3.2.5.2 Compatibility in the Literature

The standard probabilistic sense of the term "possibility" is that a possible state is one with a non-zero probability. Just one example of this view is provided by Starke from his book on automata.

> We can apply non-deterministic automata to describe the "possibilities" of a given stochastic automaton in that we call "possible" those things which have a positive probability of being turned out. [273, p. 145]

On this view "grades" of possibility are not recognized, only crisp possibility and impossibility, and thus the interpretation is that $\Pr(A) > 0 \rightarrow \Pi(A) = 1$. PPC is obviously completely in keeping with this idea, since (3.6) *equates* unitary possibility with non-zero probability.

GK-compatibility is a very strong criteria, implying both DP-compatibility and Z-compatibility (from Sec. 2.6.3.3).

**Theorem 3.9** If $p$ is GK-compatible with $\pi$ in $\Omega$, then they are Z-compatible.

**Proof:** If $p_i > 0$ then $\pi_i = 1$ so that $p_i \pi_i = p_i$. If $p_i = 0$ then $p_i \pi_i = 0$. Therefore $\sum p_i \pi_i = \sum p_i = 1$. ∎

**Corollary 3.10** If Pr is GK-compatible with $\Pi$ on $2^\Omega$ then they are DP-compatible.

**Proof:** Follows trivially from (2.119) and (3.9). ■

**Corollary 3.11** The converses of neither (3.10) nor (3.9) hold.

**Proof:** The counterexamples are situations where

$$\exists A \subseteq \Omega, \quad \Pi(A) = 1, \Pr(A) = 0, \qquad \exists \omega \in \Omega, \quad \pi(\omega) = 1, p(\omega) = 0$$

respectively. ■

Therefore for these weaker compatibility measures, (3.6) is replaced by

$$\Pr(A) > 0 \rightarrow \Pi(A) = 1, \qquad \Pr(A) = 0 \leftarrow \Pi(A) < 1, \tag{3.12}$$

which is actually an axiom of compatibility measures for Delgado and Moral [47].

Thus compatibility other than GK-compatibility significantly weakens the probabilistic identification of complete possibility with positive probability. They allow complete "possibilistic" possibility for events that are "probabilistically" impossible: having probability zero.

But remarks from some of the founders of possibility theory support PPC. Giles observes

> To assert '$A$ is impossible' is at least as strong as, and perhaps for all practical purposes equivalent to, asserting '$\neg A$'. The assertion '$A$ is possible' is more ambiguous: does the speaker mean '$A$ is *just* possible' or '$A$ is entirely possible' or something in between? The first statement implies that $A$ is unlikely; the second seems to entail no commitment at all — whether $A$ is, or is not, found to hold the speaker cannot be contradicted. [98]

Shackle recognizes that the relation between frequency and possibility is problematic.

> Plainly for every individual faced with the knowledge that out of every $n$ trials of a particular kind, a specified outcome $A$ has proved right about $m$ times, there will be some numerical value of $m/n$ below which the outcome $A$ will seem somewhat surprising or less than perfectly possible. The critical level of $m/n$ will vary with the individual and the circumstances. In this book we have little to say about the exact psychic process of forming those judgments which we are supposing to be expressed by means of an uncertainty variable. [260, p. 72]

However, not all within the fuzzy research community present a similar view. Kandel suggests the following, even stronger than PPC, although with neither justifying arguments nor references to supporting opinions.

> An important aspect of the concept of a possibility distribution is that it is nonstatistical in nature. As a consequence, if $p_Y$ is a probability distribution associated with $Y$, then the only connection between $\pi_Y$ and $p_Y$ is that impossibility (i.e., zero possibility) implies improbability but not vice versa. Thus, $\pi_Y$ cannot be inferred from $p_Y$ nor can $p_Y$ be inferred from $\pi_Y$. [134, p. 31]

Interpretation of this statement rests with the word "improbability", which can mean either $p < 1$ or $p = 0$, where $p$ is taken for $p(\omega)$ for some $\omega$. PPC (and Distribution Compatibility (3.7) in particular) is then consistent with the statement

$$\pi = 0 \rightarrow p = 0,$$

but not with

$$\pi = 0 \rightarrow p < 1,$$

although of course $p = 0 \rightarrow p < 1$.

#### 3.2.5.3 Consequences of GK-Compatibility

Note that PPC is not a definition. Probability and possibility have been well defined in Sec. 2.5.3.

Nor is it a theorem. Recall that probability is only defined on specific random sets, while possibility is only defined on consonant random sets. Thus probability and possibility are almost never even *defined* on the same random sets in order to be compared according to PPC, as shown in (2.98).

Instead, PPC is a *principle*, asserted as a *semantic* criteria. As all semantic relations, it is neither a property of nor determined by the mathematical formalism itself, as discussed in Sec. 3.1.3. It's source is methodological, outside of the specific formalisms of GIT, and intended to relate together different usages of aspects of GIT in a manner which is in accordance with these extra-theoretical considerations.

Zadeh emphasized this point when he advanced the $\gamma_Z$ compatibility measure.

> It should be understood, of course, that the possibility-probability principle is not a precise law of a relationship that is intrinsic in the concepts of possibility and probability. Rather it is an approximate formalization of the heuristic observation that a lessening of the possibility of an event tends to lessen its probability, but not vice-versa. [325]

As did Shackle.

> If *any* and *every* probability greater than zero can correspond to perfect possibility then any bi-unique mapping of potential surprise is purely arbitrary and artificial. [260, p. 113]

Since $\Pi$ and Pr are not generally defined on the same random sets, PPC interprets statements about possibility (resp. probability) in the context of a given specific (resp. consonant) random set.

For example, given the possibility distribution $\pi$ on $\mathbb{R}$ shown in Fig. 3.5 with core $\mathbf{C}(\pi) = [1, 2]$, what is the status of the expression $p(x)$? PPC allows $p$ to be any probability distribution in the class of distributions on $\mathbb{R}$

$$\{p : x \in [1, 2] \leftrightarrow p(x) > 0\},$$

effectively restricting the range of $p$ to the core $[1, 2]$. From the proof of (3.11), under Z-compatibility the condition is weaker, not *forcing* values in $[1, 2]$ to have positive probability. PPC provides no further information to determine $p$, and so by the MEP, we should choose $p$ to be uniform on $[1, 2]$.

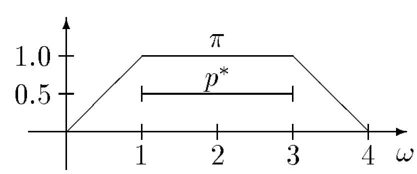


Figure 3.5: A possibility distribution and its maximum entropy GK-compatible probability distribution.

Or in the die example above, there are six possible faces, so that by PPC, for $1 \leq \omega \leq 6$, $\Pi(\{\omega\}) = 1, \Pr(\{\omega\}) > 0$, yielding distributions $\vec{p} = \langle p_i \rangle$ with $p_i > 0$ and $\vec{\pi} = \vec{\pi}^*$. Depending on the weightings of the dies, the $p_i$ have yet to be fixed, but as long as all six remain *possible*, with *some* positive probability of occurrence, the possibility values $\pi_i$ must remain at 1.

Similarly, given a Gaussian probability distribution $p(x), x \in \mathbb{R}$, what is the status of the expression $\pi(x)$? Since $\forall x \in \mathbb{R}, p(x) > 0$, by PPC, we must have the possibility distribution $\forall x \in \mathbb{R}, \pi(x) = \pi^*(x) = 1$.

The status of measure-zero events is of course an interesting one in probability theory. Whether full possibility should be allowed for such events (whether GK- or Z-compatibility should be used) is partly a methodological consideration. In the sequel the stronger GK-compatibility will be used, while noting that Z-compatibility still has the key property outlined above: a likely event must be completely possible.

Indeed, we can see that under PPC, any probability distribution with a positive value on $\forall \omega \in \Omega$ yields the maximally uninformative possibility distribution $\vec{\pi}^*$. For an $\omega$ with $p(\omega) = 0$, GK-compatibility is not more helpful, saying only that $\pi(\omega) < 1$. And Z-compatibility does not even require that.

Our conclusion is that under PPC, a standard probabilistic analysis yields essentially *no information* of a possibilistic nature. This is in keeping with our understanding of the formal relationship between probability and possibility: they are logically independent, but possibility is a much *weaker* representation of uncertainty. It is also in keeping with our common sense reasoning: when something is somewhat probable everywhere, surely it must be at least *possible* everywhere. These probabilistic cases do not imply that all *possibilistic* analyses must be so uninformative, only those with a probabilistic source, which themselves necessarily have a very *strong* informational structure.

## 3.3 Possibilistic Concepts

We can now characterize the nature of possibilistic categories, processes, and concepts. We do so here in a decidedly semi-formal manner. This discussion is intended to be suggestive, not definitive. It is intended to provide standards which guide future use of possibilistic concepts, an image of possibility which we will try to adhere to in later formal development.

Our thinking about uncertainty and indeterminism has necessarily been deeply molded by two centuries of concepts and methods which have arisen in probability theory and statistics. This is natural, since probability is the most strict representation of uncertainty. But now, given the advent of the GIT methods, some of these concepts may need to be modified, if not abandoned completely. Since this work in general, and this section in particular, is part of an attempt to break these existing mental models about how uncertain and indeterminate systems work, therefore a fair amount of time will be spent contrasting traditional stochastic with possibilistic systems.

### 3.3.1 Possibilistic Mathematics

The mathematics of possibility theory itself provide some indications of how to interpret possibilistic statements.

#### 3.3.1.1 Nonspecificity

Probability is distinguished from possibility, and indeed from all other classes of fuzzy measures on random sets, primarily in the fact that the evidence values $m_j$ are attached to singletons, and thus essentially to elements $\omega_i$. Therefore, assuming no further underlying structure in $\Omega$, probabilities can be compared only in value.

But in general, evidence values $m_j$ are attached to subsets $A_j$. It is true from Sec. 2.5.2 that the existence of a distribution $q$ establishes a structural aggregation function mapping the focal elements to the universe elements. But the evidence is still essentially valued on nonspecific subsets. And random sets with distributions are generally rare, since they have small cardinalities ($\leq n$). Therefore in general two pieces of evidences $m_1, m_2$ can not only be compared by value, but also in other ways:

- In terms of the relative cardinalities $|A_1|, |A_2|$.
- In terms of the amount of "overlap" between $A_1$ and $A_2$, measured by the relative cardinalities $|A_1 \cup A_2|, |A_1 \cap A_2|, |A_1 - A_2|$, and $|A_1 \triangle A_2|$, where $\triangle$ is the symmetric difference operator.

These additional properties are exactly what is captured by the uncertainty measures discussed in Sec. 2.6, such as nonspecificity $\mathbf{N}$, strife $\mathbf{S}$, and its special case entropy $\mathbf{H}$.

#### 3.3.1.2 Normalization

Issues of nonspecificity extend to all non-probabilistic random sets. But possibilistic random sets in particular have further restrictions. One is the normalization of (2.91), which requires a non-empty core $\mathbf{C}(\mathcal{S}) \subseteq \Omega$, whose elements are shared by all evidential claims.

Normalization places a restriction on which collections of values can be properly considered as "well-formed" statements of possibility by adhering to the requirements of possibility theory. Since mathematically, normalization is the requirement that there exists an element with unitary possibility, semantically this means that there *must* exist an element which is *completely* possible. Thus possibilistic variation posits the existence of a core of certainty around which the data set or process uncertainly varies, and which is common to all events.

Stochastic normalization also places a restriction on which collections of values can be properly considered as "well-formed" statements of probability, now by adhering to the requirements of probability theory. But by contrast, additive stochastic normalization makes the requirement that all knowledge be *accounted for* in its *division among* the various hypotheses. So stochastic variation posits the existence of a total quantity of certainty, with uncertain variation among its distinct components.

While a plausibility assignment $\vec{\mathrm{Pl}}$ which is probabilistically abnormal may be either subnormal ($\sum \mathrm{Pl}_i < 1$) or supernormal ($\sum \mathrm{Pl}_i > 1$), one which is possibilistically abnormal can only be subnormal ($\bigvee \mathrm{Pl}_i < 1$), since necessarily $\mathrm{Pl}_i \in [0,1]$. A similar normalization method can be used on either subnormal probability or possibility distributions. This is, in the possibilistic case, the dimensional extension method considered in Sec. 2.8.2, simply appending an element $\omega_{n+1}$ with $\pi(\omega_{n+1})$. In the stochastic case, the same idea results in appending an element $\omega_{n+1}$ with

$$p(\omega_{n+1}) = 1 - \sum_{i=1}^{n} p(\omega_i).$$

In either event, this method introduces a **residual hypothesis** $\omega_{n+1}$, which is appended to the distribution as a normalizing element to account for an amount of "missing" possibility or probability not present in the original distribution. Dubois and Prade [69] have remarked on this idea. Shackle [260] has described the use of this residual hypothesis in possibility theory, expressing the idea that *something* must always be completely possible, and that is the new element $\omega_{n+1}$.

In fact, Shackle generally recognized the importance of possibilistic maximal normalization, foreseeing (2.18).

> There is one constraint upon the relation between the degrees of potential surprise accorded respectively to a hypothesis and its contradictory [complement]: one or other of these degrees must be zero. For the hypothesis and its contradictory constitute between them an *exhaustive set of rival hypotheses*: everything that can happen is included by the individual under one or other of these heads. And provided that the question is a meaningful one, the individual is logically bound to suppose that there is some right answer to it. [260, p. 74]

In fact, where some modern possibility theorists do not stress the importance of possibilistic normalization, for Shackle it is of paramount importance, which "seems to us to impose itself invincibly. [260, p. 85]"

#### 3.3.1.3 Consonance and Ordinal Information

The final essential property of a fully possibilistic random set is consonance, the nesting of focal elements within each other. The core then becomes the smallest focal element with the smallest cardinality, the innermost box of a nest which spreads out from it.

This nested structure imparts a strong linear ordering to a possibilistic random set, as discussed in Sec. 2.5.3.2. This is manifested in many ways:

- In the ordering of the focal elements by inclusion $A_{i-1} \subseteq A_i$ in a nest by (2.89).
- In the ordering of the possibility values $1 = \pi_1 \geq \pi_2 \geq \cdots \geq \pi_n > 0$ by (2.92).
- In the ordering established on the elements $\omega_i$ so that $A_i = \{\omega_1, \omega_2, \ldots, \omega_i\}$ and $\pi(\omega_i) = \pi_i$ under relabeling (2.77).
- In the ordering of the membership grades of the corresponding fuzzy set $1 = \mu_{\tilde{\pi}}(\omega_1) \geq \mu_{\tilde{\pi}}(\omega_2) \geq \cdots \geq \mu_{\tilde{\pi}}(\omega_n) > 0$ as in (2.154).

Thus a discrete possibility distribution[1] can be seen as consisting of two components:

- The selection of one of the $n!$ permutations of the universe considered as a vector $\vec{\Omega} = \langle \omega_i \rangle$. The first element of the string is the core.
- The assignment of possibilistic weights to the $\omega_i$ with the only requirement that $\pi(\omega_1) = \pi_1 = 1$.

The ordering specifies a particular path through the universe, while the weights represent the "distance" in certainty values between them. This *ordinal* property of possibilistic information is crucial: we are not concerned with the division of knowledge among a set of otherwise indistinguishable entities, but rather over a set of entities which has a decidedly linear *structure*.

### 3.3.2 Possibilistic Processes

Because of nonspecificity, traditional ideas of **randomness** are completely altered in random set, and especially in possibility, theory. In probability theory we think of a given universe which is fundamentally a partition, a division of the space into distinct units. Each unit is precise and specific, and we are only uncertain as to which unit is selected, and measure the dispersion over those distinct units. Dubois and Prade comment

[1]Continuous possibility distributions are significantly more complex [230].

> Probability measures apply to precise but differentiated items of information, while possibility measures reflect imprecise but coherent items (i.e., which mutually confirm each other) ... A probabilistic model is suitable for the expression of precise but dispersed information. Once the precision is lacking, one tends to quit the domain of validity of the model. [64, pp. 6, 13]

But when thinking about random sets, or data sets or processes which are governed by general evidence theory, and in particular possibility theory, the concepts we bring to bear are very different from probability. The fundamental units, the observed events themselves, cannot in general be partitioned into additive blocks, and eventually decomposed to atomic constituents.

We can introduce the concept of a **stochastic set-process**, which is essentially an active process governed by the dynamics of random sets, rather than simple probability distributions. This is illustrated in Fig. 3.6, where for a six element universe each of the focal elements is the region labeled by its evidence value. Instead of a simple stochastic process, where individual random outcomes result, we have the image of a variety of randomly occurring subsets, with a greater or lesser degree of irreducible overlap, and with a complex, interlocking structure. The size and relative overlap of events in a stochastic set-process are constantly changing as it moves not through states $\omega \in \Omega$, but rather through *meta-states* $A \subseteq \Omega$, in a non-deterministic manner.

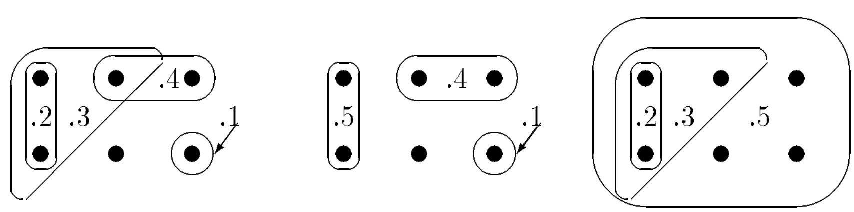


Figure 3.6: (Left) A general stochastic set-process. (Middle) Probabilistic case. (Right) Possibilistic case.

On the left of Fig. 3.6 is a generic stochastic set-process. Focal elements are related indiscriminately: disjoint, included, and properly intersecting. Because of the lack of any coherence, no distribution, and no element-subset aggregation function, is possible.

In the center is a stochastic process, here identified not on the singletons of $\Omega$, but on a disjoint class of $\Omega$. Each of these disjoint focal elements can be recognized on equivalence classes, and thus points in a lower dimensional space where $n = 4$. Nevertheless, the fundamental property of specific random sets still holds: knowledge is divided among the disjoint subsets, so that there is no ambiguity in the outcome of the stochastic set-process.

Finally on the right is a possibilistic process. Here the innermost focal element is the core, *around which* the set-process is *coherent*. In virtue of this coherence, the possibility distribution is available on the universe $\Omega$ from which the original stochastic set-process can be generated. While probability represents ambiguity, an uncertain choice among distinct alternatives, possibility represents a lack of specificity, an uncertain, but monotonically nondecreasing, distance from the central core. In a possibilistic process the core remains fixed, and the observed state varies in *extent* around it.

### 3.3.3 Statistical Interpretations

Because in a possibilistic process there is a variation not just of state, but of the *size* of the state, we have a visual image not of a point traversing a state space, but of a shifting granularity, tolerance, or precision. In statistical terms, we do not see a shifting sample mean, but rather a shifting *variance* around a *fixed* mean, the core.

It is well known that as the size of a statistical sample of a stationary stochastic process increases, the sample variance converges to a fixed value. Thus possibility theory may be appropriate to model *non-stationary* processes, where the sample variance will continue to shift around the sample mean.

Also, by the law of large numbers, large samples provide sufficient information to justify the inductive inference of taking a frequency distribution as a probability distribution, and thus to satisfy the strong constraints of stochastic representations. So possibility theory may be appropriate in modeling problems with small sample sizes. Here the weakness of the small sample is matched by the weakness of possibilistic information, while the strength of a large sample is matched by the strength of stochastic information. In fact, it might be hoped that with increased sample size, that a possibilistic analysis would become *less* useful, just as a stochastic treatment of the same problem would become *more* accurate.

Turksen has noted that probabilistic measurement depends on large numbers of observations (under classical, objectivist, probability theory, to the limit), and has suggested that possibility theory could be of some value where this was not possible.

> The implication of this view is that response data must be observed by an analyst an infinite number, or at least a large number of times for every experimental setting in a measurement study. This is generally possible in measurement experiments of physical attributes. However, it is almost an impossibility when the response data are to be provided

> by human subjects as in the case of subjective evaluations in general and measurement of fuzziness in particular ... In these cases, only a few points of response data may be observed under ordinary circumstances. Thus probabilities of fuzzy events may not in general be computed due to lack of sufficiently "large" trials. [291]

Turksen made this observation in the context of measurement of human responses, but the point holds whenever only small samples are available. But he offers no valid mathematical formulation for what possibilistic measurement might be. Instead, he simply defines the possibility value *as* the observed relative frequency from a finite number of trials. Yet this is obviously an additive distribution with an additive measure, and in no way a possibility function.

### 3.3.4 Possibilistic Locality, Extensibility, and Mutability

Possibilistic normalization provides a useful observation about the nature of possibility. While stochastic normalization is necessarily a property of the *whole* distribution $\sum_{i=1}^{n} p_i = 1$, possibilistic normalization can be satisfied by a *single element* of the distribution: $\exists \omega \in \Omega, \pi(\omega) = 1$. The existence of a normalizing element need not be unique, and in the maximally uninformative possibility distribution $\vec{\pi}^*$ *all* elements of the possibility distribution are unitary.

Thus possibility is much more of a **local** property. For example, the possibilistic normalization methods of Sec. 2.8 require only the modification or appending of a single element. However, there is a tradeoff in that after normalization a reordering of the $\pi_i$ might be necessary, in keeping with (2.92).

So possibility is also highly **mutable**: individual elements $\pi_i$ can be modified without requiring any other elements to be changed (again within reordering limits).

Similarly, possibility is highly **extensible**, since (as in the dimensional extension normalization method of Sec. 2.8.2) *any* possibility value, either unitary or non-unitary, can be appended to a possibility distribution without any requirement of global rescaling, renormalization, or recalculation (although again reordering may be required).

These properties have been recognized by Ramer and Puflea-Ramer in their description of a normalization method very similar to the focused consistent transformations of Sec. 2.141.

> Such [a] 'do-nothing' strategy may appear somewhat unusual, but it actually captures well the notion of *possibility*. Unlike probabilities, possibility values are not interrelated; for example, [the] possibility of $Y$ is

> not tied to [the] possibility of $X - Y$, unlike the complementary rule for probabilities. [232]

### 3.3.5 Possibilistic Independence from the Universe

Together, these properties obviate one of the most often criticized features of probability theory, namely its dependence on an *a priori* specified universe of discourse, or "frame of discernment". This point has been dealt with in detail by Dubois and Prade [69], and has also been stressed by Shackle.

> The circumstances in which we *can* fall back on probabilities — or at least on the orthodox classical concept of probability — are in fact severely limited. We must *first* have a clearly delimited field of possibilities or contingencies, logically exhaustive and mutually exclusive — what von Mises called a 'collective' and modern statisticians call an 'ensemble'. [260, p. 99]

The most obvious manifestation of this limitation is in the representation of ignorance, and the resulting consequences for statistical estimation. The uniform probability distribution $\vec{p}^*$ is dependent on $n$, itself a property of the universe $\Omega$, while the maximally uninformative possibility distribution $\vec{\pi}^*$ is not. When using Bayes theorem to update an *a priori* probability distribution, vastly different results will be achieved depending on what the assumed size of the universe was. Or as Dubois and Prade comment

> Probability theory offers no stable, consistent modeling of ignorance. Thus, the way a question is answered depends upon the existence of further information which may reveal more about the structure of the set of possible answers ... An important part of probabilistic modeling consists of making up a set of exhaustive, mutually exclusive alternative *before* assessing probabilities ... Quoting Cheeseman [32], 'if the problem is undefined, probability theory cannot say something useful'. [69]

The uncertainty measures $\mathbf{N}, \mathbf{S}$, and $\mathbf{T}$ from Sec. 2.6 are dependent on the cardinalities of the $A_j$. But for additive probability *itself*, independent of any *further* measure of a distribution, its requirement for an *a priori* frame reflects and results from its general dependence on the cardinalities not only of the universe, but of all the sub-events recognized within the universe. Possibility theory is free from this requirement. Shackle also discussed this.

> To sum the degrees of possibility assigned to various rival hypotheses is to fall back on the idea that it is the *number of its rivals* which gives a hypothesis its status, rather than its own particular character. [260, p. 92]

### 3.3.6 Possibility, Complexity, and Emergence

The observation in Sec. 3.3.3 about systems with small sample sizes brings us to a discussion of **complexity**. There is a great deal to say about the meaning of complexity and its relation to information (the reader is referred to at least [144,300] out of a large literature base). But one point of note is that complex systems can be said to be "historically bound". That is, they have evolved to their present states through a long series of very specific actions.

Complex systems are thus relatively impervious to traditional experimental methods. It can be very difficult, if not impossible, to establish them in initial conditions for multiple time trials. Once a measurement is made, the system can become perturbed, never to return to its previous state.

Thus complex systems generally do not yield the kind of strong time-series data required for stochastic descriptions, yet their behavior is full of uncertainty. Instead, it seems completely appropriate to attempt possibilistic analyses of such systems, which requires only weak information.

Indeed, as Weaver describes [300], stochastic methods are especially appropriate for dealing with repeated experiments on simple systems. The simplest systems are indistinguishable, as the move to Bose-Einstein counting in quantum physics requires. Here statistical techniques are paramount, as in statistical physics and thermodynamics.

"Degenerate" systems, which lack internal structure or coherence, like a mere pile of rocks or a mole of ideal gas, are best described as "aggregates". As their internal constraint and structure, and thus complexity, increases, they gain in distinguishability. In truly complex systems this can grow to be actual *uniqueness*, for example in organisms, as noted by Elsasser [82]. It is in these limit cases where repeated experiments are truly impossible.

A hallmark property of complex systems is exactly the properties of possibility, as opposed to eventuality, described by Gaines and Kohout in Sec. 3.2.5.1. By a mild abuse of language, complex systems can be described as being highly "non-ergodic". In other words, given a very large state space, only a very small portion of that space might actually be visited. This issue is discussed very clearly by Kampis [129]. Thus in such systems there will be a large number of properly possible states, while a small number of "eventual" states, perhaps even none.

In a rare, and admirable, but yet incomplete, attempt to apply fuzzy theory to a physical system, Singer relates the semantics of fuzziness directly to complexity and measurability.

> Mechanical quantities, such as distance and velocity can be measured to a relative error of $10^{-3}\%$. Electrical measuring devices such as those for current or potential allow a relative precision of 0.1%. The achievable limit in entropy and enthalpy measurements is about 1%. Even higher are the errors in chemical processes, due to autocatalysis and reaction mechanism, not defined sharply, and last, but not least, due to the fact that the velocity constants can be measured in general only indirectly, these constants are uncertain to 5% and more.
>
> In accordance with these facts, it seems justified to presuppose some unsharpness in all physical processes a priori in the sense of fuzzy set theory. This unsharpness is very small in simple mechanic [sic] and electric [sic] processes but must be taken into account in complex phenomena governed by non-equilibrium thermodynamics. [267]

Kampis suggests that this property of complex systems should actually be regarded as fundamentally **emergent**, since making a prediction in the state space usually requires (at least!) an intractable computation. Thus the occurrence of each of these rare states appears as a fundamentally *novel*, unpredictable, and thus uncertain event. It may be that the properties of possibilistic mutability and extensibility discussed above make possibility theory more appropriate to model the "surprises" such systems provide.

Possibilistic independence from the universe is also crucial here. Consider again the six-sided die from Sec. 3.2.4. When a universe of discourse can be specified (the six faces), then each member of it is identified as being completely possible, and a probabilistic analysis begins.

But instead of being simple, imagine that the die is actually a highly complex system, much of it hidden from view, which produces a number as its output. Say further that we have been led to believe that the number will be between 1 and 6, so that the *a priori* universe is as before. But in virtue of the system's complexity, a 7 may appear as an "emergent" event, causing us to extend the (known) universe. A possibilistic analysis will handle this situation gracefully, with a probabilistic one will not.

Shackle also shares in the understanding of the nature of choice in complex systems and the special inapplicability of stochastic methods to emergent systems.

> ... in all such instances [like the decision to marry or not, or the decision by a general as to a certain battle plan] the entire subsequent career of an individual or a nation is swung into one rather than another of two wholly different channels. After such a decision there can be no going back to the state of affairs which prevailed while the choice was still open. It is, accordingly, logically impossible for a person who has to make a decision in such an instance to contemplate *repeating* his experiment: such experiments are *self-destructive*. [260, p. 56]

This is, of course, common in complex physical or mechanical systems. Shackle describes this as either a lack of multiple occurrences of a single event (non-divisibility) or a lack of a population over which single occurrences are observed (non-seriability), or both.

> Isolated and, above all, self-destructive experiments, numerous enough and of crucial and dominating importance when they arise, are inherently untouchable by [probability]. [260, p. 61]

As mentioned, models of complex systems risk intractability by "combinatorial explosion". The computational efficiencies discussed above also make possibility theory attractive for complex systems modeling. Some results of possibilistic computational efficiency are described by Wierman [304].

So when reviewing classes of systems, we can place them along a dimension relating complexity, probability, and possibility. First, in simple systems (like the die, or an ergodically mixed mole of gas), beginning with complete (crisp) possibility of the states, from PPC positive probability results, in the limit a uniform distribution with average maximal probability. But in a complex system, states may be impossible, or at most properly possible, so that probability is still zero.

### 3.3.7 Capacity vs. Frequency Concepts

These observations once again emphasize that possibilistic data are not *frequency* data, since possibility values can be changed without reference to the overall sample. Possibility thus cannot be regarded in the context of other concepts which are related to frequency, such as **likelihood, chance, tendency, propensity**, or **proportion**.

Instead, possibility suggests interpretations in the context of **capacity**. Given a set of "buckets", they can all either be completely full, many can be empty, or they may be in some intermediate state, as long as at least one is full (for possibilistic normalization).

Thus possibility can be seen as relating to the general constellation of concepts surrounding capacity, including **intensity, degree of fulfillment, ease of fulfillment, distance from optimality, degree of satisfaction, similarity, resemblance, elasticity**, and **preference**. Dubois and Prade [75] in particular make this argument, and Wood and Antonson [309] use a membership function to describe a preference relation.

These concepts are all *ordinal*, measuring states by their distance from some state of maximal capacity (intensity, preference, etc.). Kosko echoes this view, but in the context of general fuzziness, not possibility.

> Fuzzy theory [is] the theory that *all things admit degrees*, but admit them deterministically. Fuzziness describes *event ambiguity*. It measures the degree to which an event occurs, not whether it occurs. Randomness describes the uncertainty of *event occurrence*. An event occurs or not, and you can bet on it. [164]

### 3.3.8 Physical Interpretations of Possibility

A point moving around a state space in a stochastic process assumes a fixed partition and a device which can observe a specific element of that partition, a specific event. We must ask what possible ontological status might attach to a *non-specific* event. Do we assume in a possibilistic process that this widening "exists" in the world, or are we just constructing models which have such variations in their observations? Is there an "actual" possibilistic process to be observed, or is there perhaps a deterministic or ordinary non-deterministic process which we are just *modeling* by possibility theory, and perhaps could have been modeled in some other formalism?

To a large extent these issues will be left as philosophical questions. But it can be seen that while the fundamental categories in probability theory are fixed, in possibility theory these categories are constantly shifting. Thus possibility can be said to be concerned with the very *definition* and *identification* of events, rather than assuming given events and then inquiring about their properties.

#### 3.3.8.1 Elastic Constraints

The term "elastic constraint" was used by Zadeh [325] to describe the effect of a "fuzzy restriction" on a variable. It is interesting to consider the possible interpretations in a *physical* system of a *physical* elastic constraint, expressed in possibilistic terms, and related to the capacity concepts described above.

Klir has suggested[2] a few potential examples. Consider a cork being pushed into a bottle. At first it can be pushed easily, then only with more force, and finally cannot be pushed further. While good mechanical models are available which describe this as a deterministic system in terms of static and dynamic friction, this simple system can be regarded *qualitatively* as well (qualitative modeling techniques will be discussed more fully in Sec. 7.1). In this case, then, there is uncertainty about the degree of force needed, or the distance the cork could be pushed in. But this uncertainty is not random: the same cork will be able to be pushed in the same distance on each trial.

As another potential example, consider a bending rod, which when bent beyond some unknown critical point $c$ will break. For a sample of rods, a frequency distribution of observed values of $c$ can be obtained, and for a large sample, a statistical model is available. But assuming that such random, probabilistic variations are small (the rods are drawn from a highly uniform stock), then the critical point is essentially deterministic, and does not vary. Again, deterministic mechanical models are available.

But as a *qualitative* problem, there is uncertainty as to the value of $c$. A natural language interpretation of the problem yields a crisp possibility distribution: before actual breakage, breakage is always possible; after breakage, it is no longer possible.

#### 3.3.8.2 Quantum Fuzziness

It is important here to mention a very large exception to our general indictment of the lack of applications of GIT methods to physical problems, and that is a growing movement to apply fuzzy logic to quantum physics [226]. It should be noted that this is a movement primarily from the physics, and not the GIT, community.

It has taken two branches. The first is in so-called "quantum logic", which uses non-distributive logics to model the causal and epistemic properties of quantum events [223, 225]. It has been suggested that Zadeh's fuzzy logics could be very useful as an alternative to the traditional probabilistic quantum logics [27, 183].

The other approach is to model quantum uncertainty not as the stochastic superposition of two crisp observables, but as a single unsharp observable [185].

#### 3.3.8.3 Other Attempts

A very few other researchers have attempted fuzzy or possibilistic models of physical systems. Cao [23] and Cao and Chen [24] use fuzzy relations to model meteorological

[2] Personal communication.

systems. And Roberts [240] applied fuzzy graph theory to the modeling of forest succession. Although not subjective evaluations, the measurement methods used by these researchers, as discussed in Sec. 3.4.4.2, is decidedly *ad-hoc*.

As mentioned in Sec. 3.3.6, Singer [267] has attempted to interpret thermodynamical systems in the context of fuzzy theory. However, he falls somewhat short of providing a sound methodological basis for the application of fuzzy theory to physical systems. For example, Onsager's law

$$I_i = \sum_{i=1}^{n} L_{ik} X_i$$

where the $I_i$ are flows, the $X_i$ forces, and the $K_{ik}$ coupling relations among them, is fuzzified according to

$$\tilde{I}_i = \bigoplus_{i=1}^{n} \tilde{L}_{ik} \odot \tilde{X}_i,$$

where $\tilde{\cdot}$ is a triangular fuzzy number and $\oplus, \odot$ are the standard addition and multiplication operators for fuzzy numbers [139]. The choice of which variables to fuzzify is driven by considerations of system complexity:

> In most cases the appropriate fuzzy expressions can be simplified, considering only a part of the variables as fuzzy. It is in all cases reasonable to consider the rate constant [of a stoichiometric equation] $k_1$ as a fuzzy quantity. $k_1$ is measured only indirectly and its 'true' value is affected in many cases by badly definable autocatalytic effects, as well. $K, c_{B_e}$ and $c_B$ [the reaction equilibrium constant, and the concentration of species $B$ at equilibrium and in general] are results of relatively precise measurements and therefore these can be regarded in most cases as crisp quantities. For reactions approaching equilibrium state slowly, $K$ and $c_{B_e}$ are fuzzy quantities as well. [267]

This analysis results in fuzzified versions of some standard versions of thermodynamic formulae, for example

$$\tilde{L} = (RT)^{-1} c_{B_e} \odot \tilde{k}_1, \qquad \tilde{X} = RT(1+K)(-1 + c_B / c_{B_e})$$

But in explaining the choice of the triangle membership function, Singer says only

> For simplifying further considerations it is assumed that the membership functions are of *triangular form*. In reality, the membership functions can in many cases be approximated with sufficient accuracy by triangles. [267]

One struggles to understand what sense of "in reality" is being used here. He is equally vague and deferring about the source of the specific memberships used, tossing off at the very end a comment which refers the reader to the standard subjective evaluation methods.

## 3.4 Traditional Semantics of Possibility

A perennial question of fuzzy systems theory is "where do the numbers come from"? While there has been a great deal of work devoted to this subject in *fuzzy* systems theory, there has been very little attention in the literature specifically to the semantics of *possibility*.[3] What attention has been paid to possibilistic semantics has been dominated by a few kinds of methods:

**Possibility as Fuzziness:** Under the interpretation of possibility as fuzziness, critiqued in Sec. 2.9.1, possibility values are determined according to the membership grades of a fuzzy set, which are themselves determined by **subjective estimation**.

**Possibility as Probability:** Under the interpretation of probability and possibility as distinct forms of information, possibility values are determined by performing a mathematical transformation of a probability distribution, itself determined by an appropriate stochastic measuring procedure.

**Possibility as Likelihood:** A small number of researchers have interpreted possibility values, *without maximal normalization*, as probabilities, but as probabilities not of a common distribution. These values are thus very similar to **likelihood** values in stochastic estimation theory.

**Objective Measurement of Fuzziness:** There are a few methods which determine membership values from objective measurement methods, and one recent method (possibilistic clustering) to objectively measure possibility distributions. However, for a variety of reasons presented below, these methods are insufficient or incomplete.

### 3.4.1 Subjective Semantics of Possibility

There is a virtual consensus in the GIT research community that fuzzy sets, and therefore, in virtue of the linkage critiqued in Sec. 2.9.1, possibility distributions,

[3] A notable exception is some of the work of Dubois and Prade [60, 64, 65, 68, 74], which helps inform our objective method presented in Chap. 4.

represent the psychological uncertainty and doubt of human subjects.

#### 3.4.1.1 Subjective Methods

There are a variety of subjective methods in the literature by which fuzzy sets, and thereby possibility distributions, are derived from heuristic principles and the opinions of people.

**Researcher Opinion** Sometimes a certain distribution is simply asserted based on the opinion of the researcher and the theoretical, methodological, or other *ad hoc* heuristic considerations which they bring to bear on the problem. For example, Giering and Kandel [97] develop a fuzzy model of resource competition among species, and introduce membership functions of optimality of a given species for a given resource. They posit that these memberships should be maximally normal (and thus possibility distributions), and then present some graphed curves with no discussion, just "Several hypothetical optimal resource membership functions ... are plotted in figure 3".

**Expert or Subject Polling** In other cases people who have expert knowledge of the modeled system are submitted to sophisticated polling techniques to provide their opinions of the possibility values. Examples are too numerous to survey even sparely. Zimmerman [328, pp. 344-349] and Hisdal [119] both provide good surveys of these "knowledge engineering" methods, and another example is the work of Hall, Szabo, and Kandel [114].

**Fuzzification** Perhaps the most prominent technique, used in virtually all fuzzy control systems, for the determination of membership values is called "fuzzification". Under fuzzification, measured crisp data are compared against a set of possibility distributions determined from some subjective method, and then aggregated to give an overall distribution of the measured data.

For example, in Fig. 3.7, three fuzzy sets, typically fuzzy numbers and thus possibility distributions, $\widetilde{F}, \widetilde{G}$, and $\widetilde{H}$, are shown. Then for a given crisp observed value $\omega \in \Omega$, a vector of values

$$\vec{\mu}(\omega) := \left\langle \mu_{\widetilde{F}}(\omega), \mu_{\widetilde{G}}(\omega), \mu_{\widetilde{H}}(\omega) \right\rangle$$

is available. In the example,

$$\vec{\mu}(x) = \langle 1, 0, 0 \rangle, \qquad \vec{\mu}(y) = \langle 0, b, a \rangle.$$

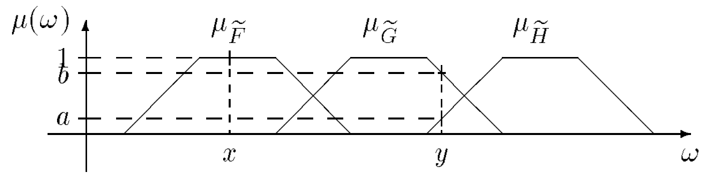


Figure 3.7: An example of fuzzification.

A typical example from the literature is the report by Gurocak and de Sam Lazaro [109]. In their fuzzy controller, the input space of the offset of a robot wrist is divided into four equally spaced fuzzy numbers labeled with the linguistic variables "very small", "small", "medium", and "big".

As a method for the determination of membership values, fuzzification is completely dependent on the fuzzy sets provided, as in the figure. Thus the question of its validity is simply begged to the question of the validity of these fuzzy sets. Typically, the number, positions, and shapes of these curves are completely *ad-hoc*, or at best, themselves determined by subjective methods.

**Neural and Genetic Training** One of the most recent subjective methods for deriving membership grades is based on training a neural network, as described by Kosko [165]. Connections among neurons in a net are modeled by weights that are usually numbers in the unit interval. The net is trained through an iterative process of being exposed to input stimuli, in the form of examples which the net is supposed to recognize. Updating occurs by strengthening connections for correct identifications, and weakening those for incorrect identifications. The net equilibrates (eventually), and the weights become stable. Methods using genetic algorithms [198] are similar in that iterative training is used.

It is natural to interpret the resulting weights as the membership grades of a fuzzy set. It may also be tempting to see this as an objective method. However, the method is crucially dependent on the selection of examples by the researchers, and perhaps more importantly on the choice by the researcher of what constitutes a correct response. The net is trained to make a distinction, but that distinction is still only a reflection of that in the mind of the trainer.

#### 3.4.1.2 Insufficiency of Subjective Methods

Even Shackle developed his possibility theory on a purely and very deliberately subjective basis.

> With us, 'possible' will mean intuitively or subjectively possible, possible in the judgment of a particular individual at a particular moment. [260,

> p. 54]

His reasoning about the nature of possibility and necessity echoes Gaines and Kohout's.

> A man cannot, in general, tell what *will* happen, but this conception of the nature of things, the nature of men and of their institutions and affairs and of the non-human world, enables him to form a judgment as to whether any suggested thing *can* happen. [260, p. 67]

But Shackle is open to criticism from a natural language argument concerning the relation between the meanings of "surprise" and "possibility" when *ontological* possibility is being considered (see Sec. 3.4.3).

Shafer is almost adamant in his *rejection* not only of additivity, but also of an objective (frequency) basis for probabilities (for him, beliefs or evidence values).

> For 250 years our culture's conception of probability has been dominated by two ideas: the idea that probabilities concerning practical matters are obtained from frequencies, and the idea that probabilities are necessarily additive. These two ideas may well have been essential to the tremendous progress that probability has made during this period. But future progress may require that we lessen our dependence on them, and this, in turn, may required that we rediscover the alternatives provided by Bernoulli and Lambert. [262]

His resolution is that additivity is not required in the context of subjective probability, but he does not consider the possibility of non-additivity in the context of *objective* probability.

> What is the meaning of the determined but not fully known probability $P$ that is supposed to lie between the bounds $p_*$ and $p^*$? If $P$ can be interpreted as a frequency or as an aleatory probability, as in Boole's work, then we can make sense of the idea that $P$ is unknown. But an unknown epistemic probability is a contradiction in terms — an unknown feature of our knowledge. Most of the recent writers on upper and lower probabilities more or less acknowledge the absence of a meaningful interpretation for the unknown additive epistemic probability $P$; they treat $P$ as a metaphor and stress that one's knowledge is fully expressed by the pair $\langle p_*, p^* \rangle$. But they still struggle to place some significance on the additivity of this metaphor, and when they try to interpret the numbers $\langle p_*, p^* \rangle$ they reveal their puzzlement as to why one's knowledge

> should fall short of an additive probability ... $p^* - p_*$, called one's 'confusion', is thought to reflect an uncertainty which somehow differs from the uncertainty reflected by additive probabilities. [262]

A few people have been troubled by the lack of an empirical basis for fuzzy sets, for example Kovalerchuk and Taliansky.

> It is very important to make progress with this old problem [of the empirical foundation of fuzzy sets] before an examination of other problems; many of them unfortunately do not have this foundation. [166]

However, almost without exception even these researchers are content to understand "empirical evidence" as the collection of the subjective evaluations of individuals. For them, subjective evaluation *is* gathering empirical data, and they do not look further to data gathered from non-human instruments.

Now this is not to say that subjective evaluations *cannot* be empirical; indeed, in many kinds of psychology the subjective evaluations of humans *are* exactly the kinds of empirical evidence required. But at the very least, subjective evaluations and objective measurements are very different *kinds* of data. Aside from methodological questions about the nature of the data gathered by psychological reports, which still trouble the foundations of psychology, the GIT community should not cut itself off from objective measurement. This view is expressed by Dubois and Prade, where they simultaneously look forward to the measurement procedure presented in Chap. 4.

> The identification of a membership function on a simple set is a problem in empirical psychometry, which is not especially difficult. But ... the expression of membership functions in terms of random sets enable statistical interpretations to live alongside pure psychometric interpretations of fuzzy sets. [70]

This issue is also discussed in Sec. 3.4.4.1.

In probability theory subjective methods also serve a useful role. But, as described by Fine [83], probability theory includes a number of *other* bases for determining probability values, including a variety of relative frequency, logical, and complexity-based methods.

Jain is one of the very few who explicitly recognizes the possibility for construction of a possibilistic (fuzzy, for him) semantics on either objective or subjective bases. He describes this in an early work on fuzzy mathematics and the construction of algebraic networks of fuzzy components.

> It should be mentioned here that the fuzzy values assigned to various components may be due to the ill-defined nature of the components, or due to subjective specification of component values. The first case is usually encountered in mechanistic systems and the second case is a common occurrence in humanistic systems. [125]

Jain does not elaborate on this idea, and does not attempt a general objective semantics for fuzziness. He does suggest that existing tolerance data could be used to assign fuzzy values. This method would be a frequency conversion method, which will be discussed below in Sec. 3.4.2.

Our purpose here is not to critique subjective methods *per se*. There can be no quarrel with their *usefulness* in GIT. They are an important component in an overall semantics of GIT, and no doubt there are situations in which they are either necessary or completely sufficient. For example, these methods are natural and useful when people control and intervene in system operation, and so psychological disposition is a serious factor. In other circumstances, there may be a good theory of the system being modeled and little or no access to physical measurement.

But the predominance of the subjectivist view for possibility has limited the areas in which possibility theory has been applied. These include applications in "knowledge" or "informational" engineering such as knowledge-based control systems, approximate reasoning, and decision support systems. In these applications the subjective semantics of possibility is appropriate, and is used to replace a human "expert" with an approximate reasoning system. Just glancing at any conference proceedings will attest to this, as well as some of the anthologies and textbooks which describe applications of GIT [64, 104, 322, 325].

In the first real textbook on GIT, Dubois and Prade mention that fuzzy theory has not yet being applied to "real" systems [55]. Not much has changed in fourteen years. Subjective methods are unsatisfactory at best for the modeling of *physical* systems or other systems in which individuals do not provide direct input. These require measurement methods which use empirically derived data, and a semantics of possibility which is not wedded to human psychology. Where possible, data should be derived from physical measurements in a manner which directly captures the possibilistic nature of that data.

#### 3.4.1.3 Predominance of Subjective Methods

The extent to which subjective semantics of fuzziness and possibility dominates GIT is truly remarkable. This view was established by Zadeh in his earliest papers, and indeed provided most of the impetus for his development of fuzzy theory: for him, *the* purpose of fuzzy theory is to model human psychology in a way that traditional information theory cannot. To that end he introduced the use of "linguistic variables" [324], which essentially use natural language terms as the values of variables.

> Much of the information on which human decisions are based is possibilistic rather than probabilistic in nature. In particular, the intrinsic fuzziness of natural languages ... is, in the main, possibilistic in origin. [325]

In this quotation we see the pillars of the Zadeh interpretation of possibility in terms of fuzzy sets: the reliance on possibility as a measure of uncertainty in human cognition for the purposes of modeling human decision making. Zadeh's linguistic calculus is necessarily purely psychological and humanistic.

> The main applications of the linguistic approach lie in the realm of humanistic systems — especially in the fields of artificial intelligence, linguistics, human decision processes, pattern recognition, psychology, law, medical diagnosis, information retrieval, economics and related areas. [324]

Hisdal suggests that the very first tenet of fuzzy theory is that

> It is possible to handle inexact information and linguistic values of variables in a mathematically well-defined way which simulates the processing of information in natural-language communication. [120]

Here can be seen the confluence of concepts in the standard fuzzy model: not only that the inexactitude of fuzzy information models the inexactitude of human psychological states, and that people represent this inexactitude in linguistic terms, but that this domain is the *only* legitimate one for fuzzy information theory. Another of her papers [119] is also steeped in the implicit view that fuzziness is subjective uncertainty, nothing more.

This view was also taken up by Sugeno in his development of fuzzy measures and integrals.

> As can be readily seen from their definition, fuzzy measures formally include probability measures as a special case. However, the concept of fuzzy measures is not used in a probabilistic environment but in a fuzzy environment where human subjectivity particularly plays an important role. [277]

Thirty years after Zadeh's introduction of the fuzzy set, a student of his echoes the identical view, specifically with regards to the "fuzzification" method for subjective evaluation.

> Fuzzification is related to the vagueness and imprecision in a natural language. It is a subjective valuation which transforms a measurement into a valuation of a subjective value, and hence it could be defined as a mapping from an observed input space to fuzzy sets in certain input universes of discourse. [174]

To my knowledge, this view has received remarkably little criticism over the years. But it must be emphasized that our argument here is not that Zadeh's view is *wrong*, only that it established unnecessary *limits* on the available interpretations of possibility.

Most researchers are fully cognizant of this marriage of subjectivity and possibility/fuzziness, and indeed they champion it in their theory and applications. Perhaps the most telling example is the entry in Singh's *Systems and Control Encyclopedia*, which defines fuzziness as being related to human language [268, p. 1862]. Even the most prominent researchers parrot the subjectivist view without any critique or analysis. Yager offers the following.

> In using either fuzzy or crisp subsets the determination of membership grades is based upon some subjective criteria of the decision maker. [315]

In a paper which introduces the purely formal concept of a fuzzy random variable, Puri and Ralescu feel compelled to justify the task in terms of a subjective theory.

> In practice we are often faced with random experiments whose outcomes are not numbers (or vectors in $\mathbb{R}^n$) but are expressed in inexact linguistic terms. [224]

Even in the introductory pages of *Possibility Theory*, Dubois and Prade assert the supremacy of subjective methods in possibility theory, without any further discussion or justification.

> Possibility functions are ... more natural for the representation of subjective information: we do not expect that a single individual will provide us with very precise data, but we would expect the greatest possible coherence in his statements. On the other hand, precise but variable data are the usual results of carefully observing a physical phenomenon. [64, p. 13]

Kandel does the same.

> Probability is an *objective* characteristic; the conclusions of probability theory can, in general, be *tested by experience*. The membership grade is *subjective*, although it is natural to assign a lower membership grade to an event that, considered from the aspect of probability, would have a lower probability of occurrence. [134, p. 74]

The pattern of identifying fuzziness with specifically linguistic variables is *almost* universal in the literature on fuzzy systems. An exception is Sugeno and Yasukawa, who carefully try to disambiguate linguistic variables from their underlying fuzzy sets.

> In an ordinary fuzzy model that is used in fuzzy control such as
>
> 1. if $x$ is positive small, then $y$ is negative small,
> 2. if $x$ is positive medium, then $y$ is positive small,
>
> then the terms "positive small", "negative small", etc., are the labels conventionally attached to fuzzy sets, where the fuzzy sets play an important role, not the labels. [278]

They go on to develop a method of inductive modeling whereby fuzzy sets which approximate observed data in a control system are first derived, and *then* a level of linguistic description is added on top of that.

Subjective methods are used even in models of *physical* systems which use fuzzy or possibility theory. For example, fuzzy dynamical systems and fuzzy differential equations can be defined as a natural generalization of their crisp counterparts [22, 161], as can fuzzy automata (as we shall see in Chap. 5). One would have imagined that such systems would be able to be applied in much of the same areas as crisp dynamical systems, for example in physics. But these systems are almost never *applied*. Instead, applications are made in systems which are regular differential equations whose *parameters* are fuzzy values, themselves determined by *subjective* methods.

An example is in the work of Kandel [132, 133], who applied these kinds of systems to subjective meteorological forecasting. In an application to mechanical systems, Sarna and Wojnarowski [253] parameterize their systems with fuzzy probability distributions. Yet even here, where there is access to stochastic information, heuristic methods are used even for determining the *probabilities* of the model.

The work of Antonson, Otto, and Wood [197, 309] is an innovative attempt to involve GIT in engineering design. But here as well, their use of fuzzy sets and possibility is limited to the subjective involvement of the human designer.

> In engineering design, stochastic (probabilistic) uncertainty won't adequately represent all uncertainties. Some uncertainties are subjective, rather than measured, for example, the coefficient of friction of a brake shoe under a variety of possible operating conditions. These subjective uncertainties can be represented by *possibility*, introduced by Zadeh. [310, p. 28]

In Blockley's otherwise exemplary application of GIT to civil engineering [14], fuzzy relations are used to represent the truly distributed nature of relations such as Miner's rule, which relates the number of stresses a body takes to the damage it sustains. Normally these are treated as crisp functions, which Blockley replaces with a fuzzy relation which he brazenly declares is derived from a subjective estimation.

Blockley justifies this approach on the grounds that he is simply making explicit the subjective judgments which are always a necessary part of modeling. While this is undoubtedly true, it is not a sufficient justification for ignoring the lack of an objective semantics. We must ask: what are the empirical sources of Miner's rule? What is the nature of the distribution of the relations in question? And how can that distribution be expressed in information theory, either stochastically, possibilistically, or fuzzily?

Kosko, on the other hand, actually *does* try to make a strong case for "ontological fuzziness", that is for using possibility as a measure of uncertainty in models of physical systems. He trumpets that "fuzziness exists", but even he makes nominalist and referential fallacies.

> Even if science had run its course and all the facts were in, a platypus would remain only roughly an mammal, a large hill only roughly a mountain, an oval squiggle only roughly an ellipse ... The only subsets of the universe that are not fuzzy are the constructs of classical mathematics. [164]

Here, in a section of his paper titled "The Universe as a Fuzzy Set", Kosko misses the point. A platypus "is" neither a mammal, a non-mammal, nor or a partial mammal independently of our (subjective) construction of the linguistic categories "animal" and "mammal". Nor are there any "subsets" in "the universe", nor "is" it one itself, any more than an electron "is" a wave-function. Adoption of this kind of mathematical realism serves only to reinforce the original Zadeh view of fuzziness, relegating it to a sophisticated method of *naming*, of *definition*, rather than as a tool for *discovery*, for *modeling*, for increasing knowledge.

### 3.4.2 Converted Probabilities

Subjective methods derive possibility values directly, either in terms of an equivalent fuzzy set or as possibility values *per se*. The other major method derives possibility distributions indirectly, working first with a probability distribution which is then transformed into a possibility distribution. None of these methods take measurements in a decidedly possibilistic manner, and they all generally violate both GK-compatibility and Z-compatibility.

#### 3.4.2.1 Frequency Distributions and Measures

In stochastic models, observations are made of the occurrence of one or another outcome $\omega_i$.

**Definition 3.13 (Frequency Distribution)** Assume a counting function $c \colon \Omega \mapsto \mathcal{W}$ such that $c_i := c(\omega_i)$ is the count of the occurrences of $\omega_i$. Then a **frequency distribution** is a function $f \colon \Omega \mapsto [0,1]$ where

$$f(\omega_i) = f_i := \frac{c_i}{\sum_i c_i}.$$

Denote the vector $\vec{f} := \langle f_i \rangle$.

**Definition 3.14 (Frequency Measure)** Given a frequency distribution $f$, then the **frequency measure** is a function $P \colon 2^\Omega \mapsto [0,1]$ where $\forall A \subseteq \Omega$,

$$P(A) := \sum_{\omega_i \in A} f_i.$$

$\vec{f}$ is a natural probability distribution with normalization $\sum_i f_i = 1$, and $P$ is a natural probability measure as in (2.83).

#### 3.4.2.2 Conversion Methods

A variety of **frequency conversion** methods are available which convert an observed frequency distribution to a possibility distribution $\vec{f} \mapsto \vec{\pi}$. Some of the more prominent ones are outlined here.

Superscripts on $\pi$ will be used to indicate the various methods. Indices on summation and maximization will be used only where necessary, and will otherwise be obvious from context.

**Maximum Normalization** From (2.32) and (2.73), the significant difference between probability and possibility is the distribution operator $\oplus$, which is $+$ for probability and $\vee$ for possibility. It is tempting, therefore, to create possibility theory directly from probability theory by simply replacing $+$ with $\vee$. As will be seen in Chap. 5, this approach has some validity.

This view can be applied to a frequency distribution (3.13) to arrive at the most common and obvious (see Klir [145, p. 104], for example) method to derive a possibility distribution.

**Definition 3.15 (Maximum Normalization Frequency Conversion)** Given a frequency distribution $f$, then let $\pi^m : \Omega \mapsto [0,1]$ be a possibility distribution where

$$\pi^m(\omega_i) = \pi_i^m := \frac{c_i}{\bigvee_i c_i}.$$

(3.15) derives possibilities directly from the counts $c \mapsto \pi^m$. It is easily seen that this is also equivalent to a frequency conversion $\vec{f} \mapsto \vec{\pi}^m$, which Klir [148] describes as a ratio scale.

**Corollary 3.16**

$$\pi_i^m = \frac{f_i}{\bigvee f_i}$$

**Proof:** Because $\sum c_i$ is a positive constant,

$$\frac{f_i}{\bigvee f_i} = \frac{c_i}{\sum c_i} \Big/ \bigvee \left( \frac{c_i}{\sum c_i} \right) = \frac{c_i}{\sum c_i} \Big/ \frac{\bigvee c_i}{\sum c_i} = \frac{c_i}{\bigvee c_i} = \pi_i^m.$$

■

The converse relation $\pi^m \mapsto f$ is given by [148]

$$f_i = \frac{\pi_i^m}{\sum \pi_i^m}. \tag{3.17}$$

The Z-compatibility of $f$ and $\pi^m$ is therefore

$$\gamma_Z(f, \pi^m) = \sum f_i \pi_i^m = \sum f_i \frac{f_i}{\bigvee f_i} = \sum \frac{f_i^2}{\bigvee f_i} = \frac{\sum f_i^2}{\bigvee f_i}$$

so that only the maximally uninformative $f$ is Z-compatible with the maximally uninformative $\pi^m$.

**Theorem 3.18** If $\gamma_Z(f, \pi^m) = 1$ then $\vec{f} = \vec{p}^*$ and $\vec{\pi}^m = \vec{\pi}^*$.

**Proof:** In general, $f_i \le \bigvee f_i$, so that

$$\frac{f_i}{\bigvee f_i} \le 1, \qquad f_i \frac{f_i}{\bigvee f_i} \le f_i, \qquad \gamma_Z = \sum f_i \frac{f_i}{\bigvee f_i} \le \sum f_i = 1.$$

In order for the equality to hold, it must be that $\forall i, f_i / \bigvee f_i = 1$ so that $\forall i, f_i = \bigvee f_i$, and therefore $\forall i_1, i_2, f_{i_1} = f_{i_2}$. Denote this shared value of all the $f_i$ as a constant $f_0$. So $\sum_{i=1}^n f_0 = n f_0 = 1$, and therefore $f_0 = 1/n$. Now, $\forall i, \pi_i^m = f_0 / \bigvee f_0 = f_0 / f_0 = 1$.

■

**Uncertainty Invariance** In a series of recent papers [95, 147, 148, 150, 152, 156, 158–160], Klir and his colleagues have developed methods for and results from using the UIP to derive a well-justified frequency conversion method which is related to maximum normalization.

Since a frequency distribution is a natural probability distribution, a frequency conversion transforms a probability distribution to a possibility distribution, and thus is a candidate for the UIP.

**Definition 3.19 (Uncertainty Invariance Frequency Conversion)** [95] Given a frequency distribution $\vec{f}$, let $\vec{\pi}^u$ be a possibility distribution such that $\mathbf{T}(\vec{f}) = \mathbf{H}(\vec{f}) = \mathbf{T}(\vec{\pi}^u)$.

**Proposition 3.20 (Log-Interval Invariance)** [95] The best justified $\vec{\pi}^u$ is

$$\pi_i^u = \left( \frac{f_i}{\bigvee f_i} \right)^a , \qquad (3.21)$$

where $a$ is solved numerically from

$$\mathbf{H}(\vec{f}) = \sum_{i=2}^{n} \left( \frac{f_i}{\bigvee f_i} \right)^a \log_2 \left( \frac{i}{i-1} \right) . \qquad (3.22)$$

Let $s := \sum_i \pi_i^{1/a}$. Then conversely, $\vec{\pi} \mapsto \vec{f}^u$ where

$$f_i^u := \frac{\pi_i^{1/a}}{s} \qquad (3.23)$$

and $a$ is solved numerically from

$$\mathbf{T}(\vec{\pi}) = -\sum_i \frac{\pi_i^{1/a}}{s} \log_2 \left( \frac{\pi_i^{1/a}}{s} \right) .$$

**Corollary 3.24** If $\mathbf{N}(\vec{\pi}^m) = \mathbf{H}(\vec{f})$, then $\vec{\pi}^u = \vec{\pi}^m$ when $\vec{\pi}^u$ is determined by log-interval scale.

**Proof:** From (3.22), (3.16), and (2.109), $a = 1$ iff

$$\mathbf{H}(\vec{f}) = \sum_{i=2}^{n} (\pi_i^m)^a \log_2 \left( \frac{i}{i-1} \right) = \mathbf{N}(\vec{\pi}^m).$$

Then from (3.21) and (3.16), $\pi_i^m = \pi_i^u$. ∎

**Necessity from Probabilistic Difference** Dubois and Prade [58] have suggested that the degree of "necessity" between two elements $\omega_{i_1}$ and $\omega_{i_2}$ is naturally expressed by the difference $f_{i_1} - f_{i_2}$, and so their "possibility" would be 1 minus that quantity. A frequency conversion method is then available.

**Definition 3.25 (Probabilistic Difference Frequency Conversion)** [58] Assume a frequency distribution $f$ ordered so that $f_i \geq f_{i+1}$, and let $f_{n+1} := 0$. Then let $\pi^d$ be a possibility distribution where

$$\pi_i^d := 1 - \sum_{j=1}^{i-1}(f_j - f_i) = i f_i + \sum_{j=i+1}^{n} f_i = \sum_{j=1}^{n} f_i \wedge f_j$$

Conversely,

$$f_i = \sum_{j=i}^{n} \frac{1}{j}(\pi_j^d - \pi_{j+1}^d).$$

This method is actually an application of the MEP as discussed in Sec. 2.6.4.2.

**Proposition 3.26** [58] Let $\mathcal{S}^{\pi^d}$ be the constructed consonant random set determined by $\pi^d$ and (2.128), with evidence function $m^d$. Then $f$ is the maximum entropy probability distribution from (2.122) of $\mathcal{S}^{\pi^d}$, so that

$$f = p^{\mathcal{S}^{\pi^d}}, \qquad f_i = \sum_{A_j \ni \omega_i} \frac{m^d(A_j)}{|A_j|}.$$

For Z-compatibility,

$$\gamma_Z(f, \pi^d) = \sum_i f_i \pi_i^d = \sum_i f_i \sum_{j=1}^{n} f_i \wedge f_j$$

which achieves unity under the same conditions as maximum normalization.

**Theorem 3.27** If $\gamma_Z(f, \pi^d) = 1$ then $f = \vec{p}^*$ and $\vec{\pi}^d = \vec{\pi}^*$.

**Proof:** Because $\sum_i f_i = 1$, therefore if $\gamma_Z(f, \pi^d) = \sum_i f_i \sum_{j=1}^n f_i \wedge f_j = 1$, it must be that $\forall f_i, \sum_{j=1}^n f_i \wedge f_j = 1$. Since $f_i \wedge f_j \leq f_i$, by the same reasoning, $\forall f_i, f_j, f_i = f_j$, and thus $\forall f_i = 1/n$. Then since $\sum_{j=1}^n f_i \wedge f_j = \sum f_i = 1$, therefore $\pi^d(\omega) = 1$. ∎

**Additivity on a Probabilistic Nest** Another observation derived from the mathematics of possibility is that, given a consonant random set $\mathcal{S} = \{\langle A_j, m_j \rangle\}$, from (2.94), each value of a possibility distribution can be expressed as the sum of some of the $m_j$. Since each of the $m_j$ is an element of a probability distribution, albeit one on $2^\Omega$, not on $\Omega$, therefore a possibility distribution "looks like" a *cumulative* probability distribution on $2^\Omega$.

Also recall from Sec. 2.9.5.4 that the alpha cuts of *any* ordered fuzzy set, even a *probability distribution*, form a consonant class on $\Omega$.

These factors are combined in another popular and common frequency conversion method of Dubois and Prade [57, 71]

**Definition 3.28 (Additive Frequency Conversion)** [57] Assume a permutation of the universe $\vec{\Omega} = \langle \omega_1, \omega_2, \ldots, \omega_n \rangle$, and let $A_i := \{\omega_1, \omega_2, \ldots, \omega_i\}$. Then given a frequency distribution $f$, let $f_{n+1} := 0$ and $\pi^c$ be a possibility distribution where

$$\pi_i^c := P(A_i) = \sum_{j=1}^{i} f_j$$

Conversely, let $\pi_0^c := 0$, then

$$f_i = \pi_i^c - \pi_{i-1}^c$$

**Corollary 3.29**

$$\forall 1 \leq i \leq n, \quad \sum_{j=1}^{i} f_j = \bigvee_{j=1}^{i} \pi_j^c$$

**Proof:** $\forall 2 \leq i \leq n$,

$$\pi_{i-1}^c = \sum_{j=1}^{i-1} f_j \leq \sum_{j=1}^{i} f_j = \pi_i^c.$$

The result follows. ■

This method does not preserve maximum uninformativeness.

**Corollary 3.30** If $\vec{f} = \vec{p}^*$ then $\forall 1 \leq i \leq n, \pi_i^c = i/n$.

**Proof:** Trivial. ■

For Z-compatibility,

$$\gamma_Z(f, \pi^c) = \sum_i \pi_i^c f_i = \sum_i f_i \sum_{j=1}^{i} f_j$$

so that $\vec{f}$ is Z-compatible with $\vec{\pi}^c$ only for the certain distribution with weight 1 on the first element of the permuted universe.

**Theorem 3.31** If $\gamma_Z(f, \pi^c) = 1$ then $\vec{f} = \vec{\pi} = \vec{1}_1$.

**Proof:** Since $\sum_{j=1}^n f_j = 1$, therefore $\forall 1 \leq i \leq n, \sum_{j=1}^i f_j \leq 1$, with equality only for $i = n$. In order for $\gamma_Z(f, \pi^c) = \sum_{i=1}^n f_i \sum_{j=1}^i f_j = 1$, it is required that $\forall 1 \leq i \leq n, \sum_{j=1}^i f_i = 1$, and so $\forall 1 \leq i \leq n, i = n$. This can only be the case if $n = 1$, and so $\vec{f} = \vec{1}_1$. $|\vec{\pi}|$ will then also be 1, with $\vec{\pi} = \vec{1}_1$. ∎

This method is ambiguous, since there are $n!$ permutations $\vec{\Omega}$. Dubois and Prade [71] suggest that $\vec{\Omega}$ should be chosen so that $f_i \leq f_{i+1}$. Then $\vec{\pi}^c$ yields the "smallest" possibility measure that contains the additive frequency measure $P$, so that $\Pi^c$ (derived from $\pi^c$ by (2.91)) and $P$ will be DP-compatible.

#### 3.4.2.3 Insufficiency of Frequency Conversions

Each of the frequency conversion methods above has its own justification and rationale. Each has advantages and disadvantages. Our concern here is not to choose among them, but rather to question the general prospect of deriving a possibility distribution based on *any* frequency distribution $\vec{f}$, especially one which yields highly Z-incompatible possibility distributions.

There can be no doubt that $\vec{f}$ is naturally an additive probability distribution. It thus generates a specific random set, denoted $\mathcal{S}^f$, with an additive probability measure $P$. So as discussed in Sec. 3.2.5.3, possibility is not even *defined* on $\mathcal{S}^f$. Instead, when we posit a frequency conversion $\vec{f} \mapsto \vec{\pi}$, we are in fact creating a *new* random set $\mathcal{S}^\pi$ based on $\mathcal{S}^f$. Such a new random set *must* be a distortion of the measured data $\vec{f}$: the data are being transformed into a form in which they do not actually exist. The measured data are maximally *specific*; the new random set is almost completely *nonspecific*.

There may in fact be a *good* conversion $\vec{f} \mapsto \vec{\pi}$. We have asserted GK-compatibility as the criteria for a good conversion, and have seen that most of the above methods satisfy Z-compatibility only for special cases. This is not surprising either. As we have seen, GK-compatibility is highly *uninformative:* every occurring event must be maximally possible, while non-occurring events can have *any* non-unitary possibility. Frequency conversion methods, on the other hand, try to provide a distinct, generally non-unitary, positive possibility for each distinct positive probability.

The problem with using frequency conversion methods is deeper, it is in $\vec{f}$ *itself*. Surely frequency conversions must be used when *only* frequency data are available. But the possibilistic representation $\vec{\pi}$ is never ultimately *appropriate* for data gathered by a frequency distribution $\vec{f}$. Instead, we should try to obtain data in a form more directly similar to the ultimate possibilistic representation. And this is what will occupy us in Chap. 4.

| $x$ | $y$ | | |
|---|---|---|---|
| | $\alpha$ | $\beta$ | $\gamma$ |
| $a$ | 0.2 | 0.1 | 0.4 |
| $b$ | 0.3 | 0.9 | 0.4 |
| $c$ | 0.5 | 0.0 | 0.2 |

Table 3.1: An example conditional probability matrix.

### 3.4.3 Possibilities as Likelihoods

Of course numbers in and of themselves have no objective "identity" as either probabilities, possibilities, or anything else. It is only a set of numbers in relation to each other which can have the properties of one kind of *distribution* or another. Statistical likelihoods share some, but not all, of the properties of probabilities, and a few researchers have suggested that they can form the basis for possibility theory.

#### 3.4.3.1 Likelihoods as Non-Additive Probabilities

Consider a conditional probability distribution presented as a matrix $\mathbf{P} = [p(x|y)]$ as shown in Table 3.1. This matrix is singly stochastic, that is $\forall y, \sum_x p(x|y) = 1$, but $\exists x, \sum_y p(x|y) \neq 1$. So when $y$ is interpreted as an independent parameter, then $p$ is a probability distribution in $x$. On consideration of the transpose $\mathbf{P}^T = [p(y|x)]$, now $y$ is regarded as the variable and $x$ the parameter. The $p(y|x)$ can still be regarded as probabilities (after all, they are derived from a table of conditional probabilities), but they are not additive in $y$ for a fixed $x$. Instead, the $p(y|x)$ are **likelihoods** denoted $L(y|x)$ in a statistical inference problem, typically that of deriving the parameter value $y$ from some measured datum $x$.

It is clear that the $L(y|x)$ form a fuzzy set, exactly as probabilities and possibilities do. And so on the Zadeh interpretation of possibility, the $L(y|x)$ can be taken as a possibility distribution. Hisdal [119], presenting her TEE method, takes this approach (in a slightly different formal context).

> Although $P(\lambda|\mu)$ denotes a probability, it is not a probability distribution *over* $\mu$, but over the different elements $[\lambda \in \Lambda]$. $P(\lambda|\mu)$ is *conditioned* on the value of $\mu$. As a function of $\mu$ (for a given $\lambda$), it is called a "likelihood distribution over $\mu$" in statistical terminology. Notice that although likelihoods are probabilities, there is no requirement that the ordinates of the likelihood function must add up to 1. However [generally] the sum of the likelihoods over all elements of $\Lambda$ must add up to 1 for each

$\mu$. [119]

The deficiency of this method is obvious: likelihoods are not possibilistic normal, although possibilistic normalization is an absolutely essential property of possibility distributions. Natvig [190], expressing a view very similar to Hisdal's, explicitly recognizes this problem, but simply dismisses out of hand any necessity for the normalization of possibility distributions.

Of course it is also true that neither possibilities nor likelihoods are additive, and this may attract people to considering likelihoods as something other than probabilities, perhaps even possibilities. And on the Zadeh interpretation of possibility, *any* fuzzy set, *including a probability distribution*, can be taken as a possibility distribution. So Natvig and Hisdal's view correctly restricts the class of possibilistic fuzzy sets to non-additive ones, but does not restrict it sufficiently to maximally normal ones.

#### 3.4.3.2 Surprise and Likelihood

In Sec. 3.4.1.2, Shackle's identification of possibility with lack of potential surprise was criticized because of its necessarily subjective basis. Instead, there are reasons to consider that "potential surprise" might be an even better description of the concept of likelihood than it is of possibility.

Recall the following quotation from Shackle.

> The occurrence of something hitherto judged impossible would cause a man a degree of surprise which is the greatest he is capable of feeling. If this be so, we have, corresponding to *perfect possibility*, a zero degree of surprise; corresponding to *impossibility*, an *absolute maximum* degree of surprise. [260, p. 68]

While these extreme cases are acceptable, the final idea that surprise is somehow *quantitatively* related to possibility cannot hold in general when considering not one's mental state with respect to an imagination of a possibility, but rather the physical possibility of an event occurring.

As an example, consider a pack of cards with 52! possible shuffles. Were I to produce a shuffle which appears well-ordered (for example, aces to kings by suit), my surprise would be very high. Indeed, among all the 52! shuffles (the specified frame), it may be maximal. Of course some event occurring outside of this frame (for example, a pink elephant appearing and eating twenty of the cards) might be even *more* surprising.

In what sense is the well-ordered shuffle less *possible* than any other shuffle? First, clearly *every* shuffle is equally *probable*. On a probabilistic analysis we can determine by the MEP that each shuffle has probability 1/52!.

Further, each shuffle is also completely *possible*. First, by PPC, because each shuffle has probability $1/52! > 0$, therefore every shuffle has possibility 1. And leaving aside PPC, on a strictly philosophical analysis, there is nothing preventing any of the shuffles from appearing.

Since all 52! shuffles are equally likely, and indeed equally, that is to say maximally, possible, therefore I *should* be just as surprised by the appearance of *any* of the 52! shuffles *if I was able to recognize them distinctly*. I am more surprised by the well-ordered shuffle because my surprise is not a function of the absolute odds of having that shuffle turn out, but rather of my cognitive capacities to recognize it as a special shuffle; I recognize it as belonging to a *subset* of shuffles which are rare with respect to all 52! shuffles. Since the most I can recognize are broad groupings of shuffles (those that look "random" vs. those that look "ordered"), then a well-ordered shuffle surprises me because within the equivalence classes of "well-ordered shuffles" vs. "not well-ordered shuffles", the well-ordered shuffle has low frequency: the *marginal* probability within the equivalence classes of a well-ordered shuffle is small compared to that of a random-appearing shuffle.

In contrast, the appearance of a 1 resulting from a roll of a six-sided die is not surprising at all because I can recognize all six faces distinctly. But in a roll of two die, snake eyes and box cars *are* surprising, although equally likely outcomes (assuming the die are distinctly identified, e.g. 3 4 is distinct from 4 3), not only because my perceptual system is biased to recognize doubles and other extreme values, but because the marginal probability of the sum of the die being 2 or 12 is low (1/36). Similarly, any *relatively* rare, but *completely possible* event results in high surprise.

So surprise is really an *epistemic* measure of *probability*, not of possibility, because it is really low *frequency* or *likelihood* events which are the most surprising. But, as Shackle goes to great length in arguing, both possibility and surprise are non-additive ("non-distributional" for him): a low degree of surprise in any one event need not affect the degree of surprise of another. So how can we move surprise from a measure of possibility to one of probability, while still relinquishing additivity? This is possible if we identify surprise with *likelihood*, which we have seen has neither stochastic nor possibilistic normalization.

It is significant to note that Shackle's argument for possibilistic normalization discussed in Sec. 3.3.1.2 is critically dependent on the assumption of a closed universe. This is despite the fact that we have seen in Sec. 3.3.5 that the lack of the requirement for a closed universe is a key advantage for mathematical possibility, and also despite Shackle's own criticism of probability being dependent on closed universes.

> A distributional [additive] uncertainty variable is peculiarly liable to misinterpretation. It assumes that the suggested answers in some finite list are all that need be taken into account for some particular question. [260, p. 51]

PPC provides some reconciliation between surprise, rare events, likelihood, and proper possibility. If we regard high surprise as low likelihood (a probability, not a possibility), then both properly possible (the pink elephant) and rare but completely possible (the shuffle) events show high surprise, since by PPC the elephant requires zero, and thus low, probability. Further, maximal surprise (zero probability) requires proper possibility. That is to say the "emergent" pink elephant, and anything else from outside the frame, would be more surprising than the simply rare ordered shuffle. While the appearance of the shuffle would be surprising, it would not be *so* surprising as to call the frame into question. *Anything* inside the frame (and thus somewhat probable, and maximally possible) is less surprising than anything outside the frame.

Further, on the interpretation of surprise as low likelihood (low probability), Shackle's criteria from the first quotation are completely consistent with PPC, depending on whether we understand them as implications or biimplications. Shackle states that perfect possibility corresponds with zero, and impossibility with maximal, surprise. By combining our interpretation of surprise and PPC, it can be derived that:

- Zero surprise implies maximal likelihood, and thus positive probability and complete possibility. But the converse does not hold: a completely possible event might have any positive probability, and thus still be somewhat surprising (like our well-ordered shuffle).

- On the other hand, impossibility implies zero probability, and thus maximal surprise. But again the converse does not hold: a maximally surprising event (like the evolution of the wing, or the Great Depression), even a zero probability event, might have any degree of (proper) possibility, and thus not be impossible.

### 3.4.4 Objective Measurement of Fuzziness

Despite the protestations above, there actually are *some* objective measurement methods to determine fuzzy set membership grades (sometimes even possibility values). However, they all have significant weaknesses and disadvantages for the objective measurement of possibility.

#### 3.4.4.1 Objective Measurement of the Subject

First, it must be recognized that there is not necessarily a clear dichotomy between "subjective" and "objective" types of measurement, and in fact there is interaction between them. Through the mediation of human action, human subjective states of mind are manifested as objective facts which are objectively observable.

So there is a real sense in which all of the subjective methods discussed in Sec. 3.4.1.1 are objective methods: they are the objective measurement of a human subject. In fact, there is a serious argument as to whether or not any other methods (for example, polling) are even possible when it is the human subject which is being measured. In other words, given that you want to construct a possibilistic model of a human operator or expert, then clearly the subjective methods are what will be used, and those will then be objective methods.

But this argument does little except move our objection to a different place. Rather than criticizing the fuzzy community for relying only on subjective methods, we must still criticize them for only modeling human subjects.

#### 3.4.4.2 Fuzzy Relations from Frequencies

Two groups of authors have advanced methods to derive two-dimensional fuzzy relations from frequency distributions. Cao [23] and Cao and Chen [24] compare a frequency distribution to itself to derive a reflexive and symmetric fuzzy relation which is possibilistic normal.

**Definition 3.32 (Cao-Chen Measurement)** [24] Given a frequency distribution $\vec{f} = \langle f_i \rangle$, let $\tilde{R} := [s_{ij}] \tilde{\subseteq} \Omega^2$ be a fuzzy relation where $\forall 1 \leq i, j \leq n$,

$$s'_{ij} = 1 - |f_i - f_j|, \qquad s_{ij} = \frac{s'_{ij} - \bigwedge_{i,j} s'_{ij}}{\bigvee_{i,j} s'_{ij} - \bigwedge_{i,j} s'_{ij}}$$

Roberts [240] compares two distinct frequency distributions by a number of *ad hoc* methods (not detailed here) to derive a general fuzzy relation.

These methods are not sufficient for our purposes, not only by the same argument as used against frequency conversions in Sec. 3.4.2, but also because they are *ad hoc* and highly specialized to produce only two-dimensional relations.

#### 3.4.4.3 Fuzzy and Possibilistic Clustering

Pattern recognition problems have been one of the primary applications of fuzzy theory for a long time. **Clustering** methods [305] have been a major method for inductively building up a pattern from measured (usually visual data). The **fuzzy** $c$**-means** method was developed early as a fuzzy method for pattern recognition. Recently more possibilistic approaches have been attempted.

**Fuzzy $c$-Means** The following definitions form the basis of the fuzzy $c$-means algorithm, and are derived from Windham [306] and Bezdek et al. [13].

**Definition 3.33 ($c$-Partition)** Let $\mathbf{X} := \{\vec{x}_k\}$ be a set of data points, each $\vec{x}_k, 1 \leq k \leq n$ a real $m$ vector. Then $\forall k$, let $U := [u_{ik}], 1 \leq i \leq c$, a $c \times n$ real matrix, be a $c$**-partition** of $\mathbf{X}$ if

$$\forall k, \quad \sum_{i=1}^{c} u_{ik} = 1, \tag{3.34}$$

$$\forall i, k, \quad u_{ik} \in [0,1], \qquad \forall i, \quad \sum_{k=1}^{n} u_{ik} > 0.$$

**Definition 3.35 (Fuzzy $c$-Means)** Let: $\vec{v} := \langle \vec{v}_i \rangle^T$ be a $c \times n$ real matrix, with each **cluster prototype** $\vec{v}_i, 1 \leq i \leq c$ a real $n$ vector; $1 \leq m < \infty$; and the **objective function** be

$$J_m(U, \vec{v}) := \sum_k \sum_i (u_{ik})^m d_{ik} \tag{3.36}$$

where $d_{ik} := |\vec{x}_k - \vec{v}_i^*|^2$ and $|\cdot|$ is an inner product induced norm.

Let $\vec{v}^*$ and $U^*$ be the **optimal prototypes** and $c$-partition of $\mathbf{X}$ respectively. The algorithm iterates the following formulae to convergence.

$$\vec{v}_i^* = \frac{\sum_i (u_{ik}^*)^m \vec{x}_k}{\sum_k (u_{ik}^*)^m}, \qquad u_{ik}^* = \begin{cases} \left( \sum_{j=1}^{c} (d_{ik}/d_{jk})^{\frac{1}{m-1}} \right)^{-1}, & \forall i, d_{ik} > 0 \\ 0, & \exists i, d_{ik} = 0 \end{cases}.$$

**Possibilistic Clustering** Note the odd additivity requirement of (3.34). The rows of $U$ are hypothesized to be fuzzy sets, but the columns are conditional probabilities. In fact, while $U$ is a fuzzy relation, it is also a stochastic matrix (recalling that *all* stochastic matrices are fuzzy relations).

To move fuzzy $c$-means towards possibilistic clustering, it seems reasonable to replace the requirement (3.34) with

$$\forall k, \quad \bigvee_{i=1}^{c} u_{ik} = 1, \tag{3.37}$$

making $U$ a possibilistic matrix (matrix of conditional possibilities, see Sec. 5.3.2.4).

Krishnapuram and Keller [170] have suggested a possibilistic modification of fuzzy $c$-means, replacing (3.36) with

$$J_m(U, \vec{v}) := \sum_k \sum_i (u_{ik})^m d_{ik} + \nu_i (1 - u_{ik})^m,$$

where $\nu_i$ can be determined in a variety of ways, and is generally on the order of $d_{ik}$.

Instead of (3.37), they require

$$\forall k, \quad \bigvee_{i=1}^{c} u_{ik} > 0,$$

The possibilistic normalization of (3.37) is dependent on a prototype being a data point, although this can be forced by choosing $\vec{v}_i$ to be the $\vec{x}_k$ closest to the cluster center (mean). Thus their approach is flawed from the perspective of mathematical possibility theory.

Another possibilistic clustering approach which shows some merit is the so-called "mountain method" of Barone et al.

**Definition 3.38 (Mountain Function)** [9] Let: $\mathbf{X} := \{\vec{x}_k\}$ be a set of **data points**, each $\vec{x}_k, 1 \le k \le n$ a real $m$ vector; $\mathbf{F} := \{\vec{f}_l\}$ be a set of **grid points**, each $\vec{f}_l$ a real $m$ vector; $d(\vec{x}_k, \vec{f}_l)$ be a **distance** between $\vec{x}_k$ and $\vec{f}_l$; and $\alpha > 0$ arbitrary. Then $\forall \vec{f}_l$, the **mountain distribution** is the possibility distribution

$$\pi_l(\vec{x}_k) := e^{-\alpha d(\vec{x}_k, \vec{f}_l)},$$

and the **mountain function** is

$$M(\pi_l) := \sum_{k=1}^{n} \pi_l(\vec{x}_k).$$

The algorithm proceeds by first identifying $\max_{\vec{f}_l} M(\pi_l)$, and hypothesizing it as the first cluster center $\vec{v}_l$, then deleting $\vec{v}_l$ from $\mathbf{F}$, adjusting the $\vec{f}_l$ to reflect the loss, and then iterating.

The function $\pi_l(\vec{x}_k)$ indicates the possibility that $\vec{f}_l$ is a point contained in the same cluster as $\vec{x}_k$, and $M(\pi_l)$ serves as an information measure of $\vec{f}_l$. It is thus useful to consider this as a problem of data induction, and apply the MUP by replacing $M(\pi_l)$ with $\mathbf{N}(\pi_l)$.

# Chapter 4

# Possibilistic Measurement

A concept without a percept is empty; a
percept without a concept is blind.

— *Kant*

To the extent that GIT researchers are satisfied with the more modest goals of "informational engineering", then the nominal Zadeh program, where fuzzy sets which more or less reflect our natural language definitions are constructed *ad hoc*, may be sufficient. But a real "ontological" program for possibility must moreover try to find *good* definitions specifically in accordance with possibilistic mathematics. Acceptable definitions must be *constrained* by something *in the world*, and something other than our will and whim. This is the essence of empiricism, to open our eyes and only to construct our theories to be consistent with what is *seen*; and beyond that, to learn to see new things and thus extend our categories.

From the arguments of Secs. 3.1 and 3.4.1, possibilistic *measurement procedures*, methods of *empirical observation*, are required. In Sec. 3.4.2 it was shown that probabilistic measurement procedures which yield frequency distributions are not sufficient. Such data are necessarily specific, and thus not appropriate for possibilistic representations. Specific data have *very* strong informational structures, much stronger than the very weak possibilistic structures. The difference between stochastic and possibilistic information, as expressed by PPC, is extraordinary: probability distributions provide virtually no possibilistic information, and thus virtually all conversions from frequency distributions yield incompatible possibility distributions.

Instead measurement procedures are required that yield data in accordance with the semantic aspects of possibility theory outlined in Secs. 3.2 and 3.3, and thus governed by possibility measures and distributions. In particular, *non-specific* data are necessary, which do *not* yield traditional frequency distributions.

## 4.1 Measuring Devices

As discussed in Sec. 3.1.2.2, measurement is the general process of encoding an aspect of the "real world" into its representation in a formal system. It is only through measurement procedures that we can gain knowledge about the world; it is through the results of measurements that the world is "presented" to us. In our case, this formal system is the universe of discourse, the set $\Omega = \{\omega\}$. For the moment, the structure of $\Omega$ is not specified.

### 4.1.1 Physical Measuring Devices

We generally think of a measuring device as producing a measured value which is a number $\omega \in \Omega$. For example, a thermometer calibrated in integral degrees in the interval $[0, 100]$ would yield a result, say 72 degrees, $72 \in \{0, 1, \ldots, 100\}$.

On closer examination, however, we recognize that there is uncertainty on the readout of the thermometer. The thermometer is in fact a glass tube, a continuous object, which we will represent as $\Omega \subseteq \mathbb{R}$. The tube is marked at certain points, say $d_j$, marked with a certain number of degrees. When the thermometer equilibrates, the mercury stops at some point almost always *between* two of the marked points.

While we can use subjective estimation to interpolate between these two points, within the formalism (or for a digital, electronic thermometer) only an *interval*, say $[d_j, d_{j+1})$ denoted as $B_j$, can be reported as the result of the measurement. While any particular interval $B_j$ is usually identified by and reported as a single number, either $d_j$ or $d_{j+1}$, it must always be kept in mind that it in fact indicates the *entire* interval $[d_j, d_{j+1})$. Observation of a specific position of the mercury (an $\omega \in B_j$) must yield an interval readout $B_j$. Thus observation of only the interval $B_j$ leaves uncertainty as to the "actual" value $\omega \in B_j$.

### 4.1.2 General Measuring Devices

The following definition is used in order to incorporate such uncertainty into measuring devices.

**Definition 4.1 (General Measuring Device)** A **general measuring device** is a class

$$\mathcal{C} := \{A_{j'}\} \subseteq 2^{\Omega}, \qquad 1 \leq j' \leq N' := |\mathcal{C}|,$$

where each $A_{j'} \subseteq \Omega$ is called a **measurable set.**

$\mathcal{C}$ is the collection of all subsets which are observable by the device. Each time a measurement is taken, an $A_{j'} \subseteq \Omega$ results as a report of $\mathcal{C}$. The nature of the measuring device will depend on the elements and structure of $\mathcal{C}$. In the thermometer example, $\mathcal{C} = \{B_j\}$ is the collection of disjoint, equal length, half-open intervals $B_j = [d_j, d_{j+1})$.

This usage is in keeping with the representations of "events" as subsets in mathematical measure theory, both classical [115] and fuzzy [299]. $\mathcal{C}$ as a measuring device is also a direct generalization of the standard sense of a point measuring device from Sec. 4.1.1, since if $\forall A_{j'} \in \mathcal{C}, \exists \omega \in \Omega, A_{j'} = \{\omega\}$, then the number-reporting device is recovered.

**Definition 4.2 (General Measurement Record)** Assume a general measuring device $\mathcal{C}$. Let each distinct measurement result be indexed by $s$ for $1 \leq s \leq M$, and denote observation $s$ from $\mathcal{C}$ as $A^s$, so that $\forall s, \exists! j', A^s = A_{j'}$. Then a **general measurement record** is a vector

$$\vec{A} := \langle A^s \rangle = \left\langle A^1, A^2, \ldots, A^M \right\rangle .$$

### 4.1.3 Empirical Random Sets

Since $\vec{A}$ is a vector, it may be that $\exists s_1, s_2, A^{s_1} = A^{s_2}$.

**Definition 4.3 (Empirical Focal Set)** Given a general measuring device $\mathcal{C}$ with measurement record $\vec{A}$, let

$$\mathcal{F}^E := \{A_j\} = \{A_1, A_2, \ldots, A_N\}$$

be an **empirical focal set** derived by eliminating the duplicates from $\vec{A}$, where:

$$1 \leq j \leq N, \qquad \mathcal{F}^E \subseteq \mathcal{C}, \qquad N \leq N', \qquad N \leq M, \qquad \forall A_j \in \mathcal{F}^E, \exists A^s \in \vec{A}, A^s = A_j,$$

and inclusion of an element in a vector is defined as appropriate.

Each of the $A_{j'}$ is one of the $N'$ measurable sets; each of the $M$ sets $A^s$ is a record that one of the $A_{j'}$ has been observed; and finally each of the $A_j$ is one of the $N$ sets which was actually observed at one time or another.

Now proceed in a manner analogous to frequency distributions (3.13).

**Definition 4.4 (Set-Frequency Distribution)** Given a general measurement record $\vec{A}$ and empirical focal set $\mathcal{F}^E$, then let $C : \mathcal{F}^E \mapsto \mathcal{W}$ be a **set counting function**, where $\forall A_j \in \mathcal{F}^E, C_j := C(A_j)$ is the number of occurrences of $A_j$ in $\vec{A}$. Then a **set-frequency distribution** is a function $m^E : \mathcal{F}^E \mapsto [0,1]$ where

$$m^E(A_j) := \frac{C_j}{\sum_{A_j \in \mathcal{F}^E} C_j} = \frac{C_j}{M}, \qquad m^E_j := m^E(A_j).$$

**Corollary 4.5** $m^E$ is an evidence function.

**Proof:** Since $\sum_j C_j = M$, therefore $\sum_j m_j^E = \sum_j C_j/M = 1$. The result follows from the definition of evidence functions (2.52). ∎

**Definition 4.6 (Empirical Random and Focal Sets)** Given a set-frequency function $m^E$, let the **empirical focal set** $\mathcal{S}^E$ be the random set derived as in the definition (2.62), with **empirical focal set** $\mathcal{F}^E$.

Set-based statistics and empirically derived random sets have only a small presence in the literature. Within GIT they have been used primarily by Wang and Liu [296, 297] and Dubois and Prade [65, 68, 74]. Fung and Chong [90] provide an interesting example of the use of set statistics in their critique of Dempster's rule of combination (however, their argument is flawed, and there are errors in a crucial table).

It should also be noted that there is an entirely different sense of "set-valued statistic", as used by Degroot and Eddy [46]. Here it does not mean a mathematical property of some measured subset data, but rather an indeterminate value for the parameter of a probability distribution. For example, a uniform probability distribution would have a set-valued parameter if it was defined on a disconnected subset of the line, for example $[1,2] \cup [5,6]$.

### 4.1.4 Disjoint Measuring Devices

The key feature of a classical instrument described in Sec. 4.1.1 is that its measurable sets are *disjoint*.

**Definition 4.7 (Disjoint Measuring Device)** A general measuring device $\mathcal{C}$ is a **disjoint measuring device** if $\forall A_1, A_2 \in \mathcal{C}, A_1 \perp A_2$.

Generally, scientists strive to construct disjoint measuring devices. In such devices $\mathcal{C}$ is an equivalence class on $\Omega$, establishing observations of $\omega \in \Omega$ in an equivalence relation, and yielding the observations $A^s \in \mathcal{C}$ unambiguous. Because the $A_j$ are disjoint, observation of any one particular subset admits to no uncertainty *at the level of description of* $\mathcal{C}$.

Furthermore, when $\mathcal{C}$ is a partition, that is $\bigcup_{j'=1}^{N'} A_{j'} = \Omega$, then $\mathcal{C}$ covers $\Omega$, yielding all observations possible. Alternatively, even when $\mathcal{C}$ does not cover $\Omega$, $\mathcal{C}$ does cover the sub-universe $\left(\bigcup_{j'=1}^{N'} A_{j'}\right) \subseteq \Omega$. Measurement of a given $A_{j'} \in \mathcal{C}$ leaves us with uncertainty about the value $\omega \in A_{j'}$. The cardinalities $|A_{j'}|$ relative to $|\Omega|$ indicates the precision of the thermometer.

So in this case $\mathcal{C}$ can itself be considered as a *new* universe of discourse $\Omega' := \mathcal{C} = \{A_{j'}\}$. Of course $\Omega'$ is essentially equivalent to $\Omega$. The difference is just that in $\Omega'$ the $\omega$ are grouped into the sets $A_{j'}$, and $\Omega'$ is considered as a collection of the $A_{j'}$, not of the $\omega$.

Because the $A_{j'}$ are disjoint, so will the actual observed subsets $A_j$. Then $\vec{A}$ becomes a time-series data set on points in $\Omega'$, and the empirical evidence function $m^E$ becomes a frequency distribution over the disjoint $A_{j'}$, as a true probability distribution, and not as an evidence function.

**Proposition 4.8** If $\mathcal{C}$ is a disjoint measuring device, then as in Sec. 3.13, let

$$c' \colon \Omega' \mapsto \mathcal{W} \qquad f' \colon \Omega' \mapsto [0,1], \qquad P' \colon 2^{\Omega'} \mapsto [0,1]$$

be a point counting function, frequency distribution, and frequency measure respectively, where $\forall 1 \leq j' \leq N'$ and $\forall B \subseteq \Omega'$,

$$c'(A_{j'}) := C(A_{j'}), \qquad f'(A_{j'}) := c'(A'_j)/M, \qquad \sum_{j'} c'(A_{j'}) = 1, \qquad P'(B) = \sum_{A_{j'} \in B} f'(A_{j'}).$$

Thus time-series data on classical instruments necessarily generate probability distributions. As argued in Sec. 3.4.2.3, a frequency conversion $f' \mapsto \pi$ can be constructed, but it is better to continue the search for appropriately possibilistic measured data.

### 4.1.5 Incomplete Observations

Consider a classical measuring device which yields counts (here denoted simply $c_i := c(\omega_i)$), and assume that a certain observation is missing from the data set. Such a missing observation is anywhere in $\Omega$, and the count of such missing observations can be denoted as $c_0$ with frequency $f_0$.

The frequency $f_0$ can be regarded as an observation of the completely nonspecific subset $\Omega$, and thus the point data stream, together with its missing observations, can be regarded as resulting from measurements on a special device.

**Definition 4.9 (Augmented Specific Measuring Device)** Given $\Omega$, let

$$\mathcal{C}^+ := \{\{\omega_1\}, \{\omega_2\}, \ldots, \{\omega_n\}, \Omega\}$$

be the **augmented specific measuring device**.

$\mathcal{C}^+$ is *almost* completely specific, consisting of the singleton sets together with the whole universe.

**Corollary 4.10** If $\mathcal{C} = \mathcal{C}^+$, then $\forall 1 \leq i \leq n, \mathrm{Pl}_i = f_i + f_0$, where $f_i$ is the frequency of $\omega_i$, $f_0$ is the frequency of $\Omega$, and Pl is the plausibility on $\mathcal{S}^E$.

**Proof:** From the plausibility assignment formula (2.68),

$$\mathrm{Pl}_i = \sum_{A_j \ni \omega_i} m_j^E = m^E(\{\omega_i\}) + m^E(\Omega) = f_i + f_0.$$

∎

For a simple example (illustrated in Fig. 4.1), assume that a system with three states $\Omega = \{a, b, c\}$ is observed at ten uniformly distributed times, with $a$ and $c$ each seen twice, $b$ seen three times, and three cases where the sensor made no report. In the augmented specific device, these final three cases are recorded as observations of $\Omega$, and the specific observations replaced with observations of the singleton sets $\{a\}, \{b\}$, and $\{c\}$ respectively. The overall empirical random set is then

$$\mathcal{S}^E = \{\langle\{a\}, 1/5\rangle, \langle\{b\}, 3/10\rangle, \langle\{c\}, 1/5\rangle, \langle\Omega, 3/10\rangle\}$$

with plausibility assignment $\vec{\mathrm{Pl}} = \langle 1/2, 3/5, 1/2\rangle$, which is neither an additive probability distribution nor a maximal possibility distribution.

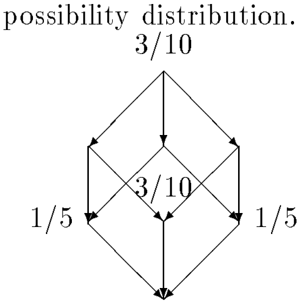


Figure 4.1: An augmented specific random set for incomplete point observations.

Other researchers have considered this issue. In some sense the observation of $\Omega$ acts as the "residual hypothesis" as used by Shackle, and as discussed in Sec. 3.3.1.2. Shafer [262] has discussed Lambert's use of a ternary calculus valued on the amount $a$ of support for a proposition, that amount $e$ opposed, and that amount $u$ which is neutral. Shafer has shown that Lambert's results are equivalent to a random set from an augmented specific device on a two-dimensional universe, such that $a$ and $e$ value the singletons, and $u$ the universe ($u$ being neither $a$ nor not-$a = e$). Bogler [17] discusses observations of the universe in the context of evidence theoretic measurement, and Klir [146] describes the special case with two binary variables. Dubois and Prade [69] consider the maximum entropy distribution on $\{A, B, \Omega - \{A, B\}\}$. And Gertler and Anderson [96] consider Dempster's combination rule with mixed singleton/universe focal sets.

### 4.1.6 Continuous and Discrete Spaces

In the earlier chapters $\Omega$ has been a finite space. But as we move to discuss possibilistic measurement, it will be desirable to let $\Omega = \mathbb{R}$, and the measurable subsets $A_{j'}$ be closed intervals in $\mathbb{R}$. Mathematical complications can be avoided as long as $\mathcal{S}^E$ is finite, that is as long as only finitely many observations $N$ are taken. This is because an interval $A = [a, b] \subseteq \mathbb{R}$ can be characterized completely by the two endpoints $a$ and $b$. With each new observation, $N$ grows by at most 1, and so the number of endpoints grows by at most 2. It is only these endpoints that need to be recorded, and none of the properties of the continuum of points between them is significant. Therefore finite random sets on $\mathbb{R}$ can be represented as random sets on the finite space of the set of all these interval endpoints. This will be completely discussed below in Sec. 4.2.1.

## 4.2 Possibilistic Histograms

Possibility distributions derived from consistent empirical random sets can be properly described as **possibilistic histograms**, similar to ordinary (stochastic) histograms, but resulting from overlapping interval observations, and thus governed by the mathematics of random sets.

**Definition 4.11 (Possibilistic Histogram)** Assume $\mathcal{S}^E$ is consistent, or is a consistent approximation. Then a **possibilistic histogram** is the possibility distribution $\pi$ determined from the plausibility assignment formula (2.68).

**Corollary 4.12 (Possibilistic Histogram Formula)** If $\pi$ is a possibilistic histogram, then $\forall \omega \in \mathbf{U}(\pi)$,

$$\pi(\omega) = \sum_{A_j \ni \omega} m_j^E = \frac{\sum_{A_j \ni \omega} C_j}{M}.$$

**Proof:** Follows immediately from the plausibility assignment formula (2.68) and the set-frequency function distribution definition (4.4). ∎

### 4.2.1 The Form of Possibilistic Histograms

In order to analyze the properties of possibilistic histograms it is necessary to mathematically describe their components. The following definitions are summarized in Table 4.1, and are illustrated in the example in Sec. 4.2.2 and in Fig. 4.2.

**Definition 4.13 (Empirical Focal Set Components)** Let $\Omega = \mathbb{R}$, and assume a random set $\mathcal{S}^E$.

- Let each observed subset $A_j \in \mathcal{F}^E$ be a closed interval denoted by its endpoints $A_j := [l_j, r_j]$.
- Let $l_{(j)}$ and $r_{(j)}$ be the **order** and **"reverse order"** statistics [44] of the left and right endpoints, so that

$$l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(N)}, \qquad r_{(N)} \leq r_{(N-1)} \leq \cdots \leq r_{(1)}. \tag{4.14}$$

  are permutations of the $l_j, r_j$.

- Denote the vectors of **endpoints** and **ordered endpoints** as

$$\vec{E}^l := \langle l_1, l_2, \ldots, l_N \rangle, \qquad \vec{E}^r := \langle r_1, r_2, \ldots, r_N \rangle, \qquad \vec{E} := \langle l_1, l_2, \ldots, l_N, r_1, r_2, \ldots, r_N \rangle .$$

$$\hat{E} := \left\langle l_{(1)}, l_{(2)}, \ldots, l_{(N)}, r_{(N)}, r_{(N-1)}, \ldots, r_{(1)} \right\rangle .$$

**Theorem 4.15 (Consistent Endpoints)** If $\mathcal{F}^E$ is consistent then

$$\max_j l_j = l_{(N)} \leq r_{(N)} = \min_j r_j,$$

so that $\mathbf{C}(\pi) = [l_{(N)}, r_{(N)}]$.

**Proof:** Let $\mathbf{C}(\pi) = [l, r]$, and assume $l_{j_1} < l_{j_2}$. Then $\forall \omega \in [l_{j_1}, l_{j_2}), \omega \notin A_2$. So since $\forall \omega \in \mathbf{C}(\pi)$, therefore $\nexists l_j < l$, and so $l \geq \max l_j = l_{(N)}$. If the inequality is strict, then $\exists \omega \in (l_{(N)}, l)$, which cannot be, since $l_{(N)}$ is the leftmost left endpoint, and there are no other $A_j$ available for $\omega$ to be a member of. Therefore the equality holds, and $l = l_{(N)}$. The result $r = \min r_j = r_{(N)}$ follows by an analogous argument. Finally, $l_{(N)} = l \leq r = r_{(N)}$. ■

Note that if $l_{(N)} = r_{(N)}$ then $\pi$ has a point core.

**Corollary 4.16 (Endpoint Ordering)** If $\mathcal{F}^E$ is consistent then the joint linear order on $\hat{E}$ is

$$l_{(1)} \leq l_{(2)} \leq \cdots \leq l_{(N)} \leq r_{(N)} \leq r_{(N-1)} \leq \cdots \leq r_{(1)}.$$

**Proof:** Trivial from the definition (4.13) and consistent endpoint conditions (4.15). ■

**Definition 4.17 (Possibilistic Histogram Components)**

- Let

$$E := \{e_k\}, \qquad E^l := \{e^l_{k^l}\}, \qquad E^r := \{e^r_{k^r}\}$$

be the sets of endpoints with duplicates omitted from $\vec{E}, \vec{E}^l$ and $\vec{E}^r$ respectively, ordered as for endpoints (4.16), where

$$\forall e_k \in \vec{E}, \qquad \forall e^r_{k^r} \in \vec{E}^r, \qquad \forall e^l_{k^l} \in \vec{E}^l,$$

$$1 \le k \le Q := |E|, \qquad 1 \le k^l \le Q^l := |E^l|, \qquad Q^r := |E^r| \ge k^r \ge 1,$$

so that $E = E^l \cup E^r$ and $Q^l + Q^r = Q$.

- Let
$$G_k := \begin{cases} [e_k, e_{k+1}), & e_k, e_{k+1} \in E^l \\ [e_k, e_{k+1}], & e_k \in E^l, e_{k+1} \in E^r \\ (e_k, e_{k+1}], & e_k, e_{k+1} \in E^r \end{cases}$$
for $1 \le k \le Q - 1$.

- Let
$$T_k := \{\langle x, y\rangle \in \mathbb{R} \times [0,1] : x \in G_k, y = \pi(x)\}.$$
for $1 \le k \le Q - 1$.

- For an (open or closed) interval $I \subseteq \mathbb{R}$ and $y \in [0,1]$, let $\pi(I) = y$ denote that $\forall x \in I, \pi(x) = y$.

**Theorem 4.18** $N + 1 \le Q \le 2N$

**Proof:** The inequalities in (4.14) will be strict or not depending on whether a pair $A_{j_1}, A_{j_2}$ share an endpoint. All the $A_j$ are distinct, so they cannot share both endpoints. This forces most, but not all, of the $l_j, r_j$ to be distinct. Consider first a single observation $A_1 := [a, b]$. When $a = b$ then $A_1$ is a point observation. When a second observation $A_2 := [c, d]$ is made, then there are four possibilities:

$$c = d \in \{a, b\}, \qquad c \in \{a, b\}, d \notin \{a, b\}, \qquad c \notin \{a, b\}, d \in \{a, b\}, \qquad c \notin \{a, b\}, d \notin \{a, b\}.$$

As distinct, consistent, observed intervals are added, in one limit all the $l_j, r_j$ are distinct, so that $Q = 2N$; in the other they all share only a common point core $r_{(N)} = l_{(N)}$, so that $Q = N + 1$. ■

Each of the $e_k$ is equal to at least one of the (left or right) observed endpoints $l_j$ or $r_j$, and $\pi$ is completely determined by the coordinates $\langle e_k, \pi(e_k)\rangle$. $\pi$ is piecewise constant, consisting of the intervals $T_k$. Each $e_k$ marks a discrete jump either up to $\pi(e_k)$ or down to $\pi(e_k + 1)$, depending on whether $e_k \in E^r$ or $e_k \in E^l$.

**Theorem 4.19 (Possibilistic Histogram Form)** If $\pi$ is a possibilistic histogram, then:

1. $\mathbf{C}(\pi) = [e^l_{Q^l}, e^r_{Q^r}]$.
2. $\mathbf{U}(\pi) = [e^l_1, e^r_1] = \bigcup_{k=1}^{Q-1} G_k$.
3. $\pi([-\infty, e^l_1)) = \pi((e^r_1, \infty]) = 0$.
4. $\forall e_k \in E^l, \pi(G_k) = \pi(e_k)$.
5. $\forall e_k \in E^r, \pi(G_{k-1}) = \pi(e_k)$.

**Proof:**

1. Since $e^l_{Q^l} = l_{(N)}$ and $e^r_{Q^r} = r_{(N)}$, this follows from consistent endpoint properties (4.15).
2. From the endpoint ordering (4.16), $e_{Q^l+1} = e_{Q^r}$ and $e_{Q^l+2} = e_{Q^r-1}$, and for $1 \le k \le Q^r, e_{Q^l+k} = e_{Q^r-k+1}$. Therefore

$$\begin{aligned}
\bigcup_{k=1}^{Q-1} G_k &= \left(\bigcup_{k=1}^{Q^l-1} G_k\right) \cup G_{Q^l} \cup \left(\bigcup_{k=Q^l+1}^{Q-1} G_k\right) \\
&= [e_1, e_2) \cup [e_2, e_3) \cup \cdots \cup [e_{Q^l-1}, e_{Q^l}) \cup [e_{Q^l}, e_{Q^r}] \cup \\
&\qquad (e_{Q^r-1}, e_{Q^r-2}] \cup \cdots \cup (e_{Q-1}, e_Q] \\
&= [e_1, e_Q] = [e^l_1, e^r_1] = [l_1, r_1] = \cup_{j=1}^{N} A_j = \mathbf{U}(\pi).
\end{aligned}$$

3. Follows immediately from (2) and the definition of support (2.39).
4. Fix $e_k \in E^l$, and fix $x \in G_k = [e_k, e_{k+1})$. Since there is no $A_j$ for which $e_k < l_j < e_{k+1}$ or $e_k < r_j < e_{k+1}$, therefore

$$\forall x_1, x_2 \in G_k, \quad \{A_j : x_1 \in A_j\} = \{A_j : x_2 \in A_j\}.$$

   The result follows from $e_k \in G_k$.
5. Follows from an analogous argument to (4).

■

### 4.2.2 An Example

Fig. 4.2 shows an example of a possibilistic histogram for $\Omega = [1, 5]$ and the measurable set $\mathcal{C}$ is the Borel field of $\Omega$. Four subset measurements are made yielding the measurement record

$$\vec{A} = \langle [1.5, 3.5), [1, 2), [1, 2), [1.5, 4) \rangle .$$

| Group | Components | Bound | Description |
|---|---|---|---|
| $\mathcal{C}$ | $A_{j'}$ | $N'$ | Measurable class |
| $\vec{A}$ | $A^s$ | $M$ | Measurement record |
| $\mathcal{F}^E$ | $A_j$ | $N$ | Empirical focal set |
| $\vec{E}$ | $l_j, r_j$ | $2N$ | Endpoints vector |
| $\vec{E}^l, \vec{E}^r$ | $l_j, r_k$ | $N$ | Left and right endpoints vectors |
| $\hat{E}$ | $l_{(j)}, r_{(j)}$ | $2N$ | Ordered endpoints vector |
| $E$ | $e_k$ | $Q$ | Endpoints set |
| $E^l, E^r$ | $e^l_{k^l}, e^r_{k^r}$ | $Q^l, Q^r$ | Left and right endpoints |
| | $G_k$ | $Q-1$ | Domain interval of $\pi$ |
| | $T_k$ | $Q-1$ | Function interval of $\pi$ |

$$N \leq N', \qquad N \leq M, \qquad N+1 \leq Q \leq 2N, \qquad Q = Q^l + Q^r$$

Table 4.1: Summary of components of possibilistic histogram $\pi$.

After eliminating duplicates, then the set of observed intervals $\mathcal{F}^E$ with $N = 3 < M = 4$ and random set $\mathcal{S}^E$ are

$$\mathcal{F}^E = \{[1,2), [1.5, 3.5), [1.5, 4)\}, \qquad \mathcal{S}^E = \{\langle [1,2), .5\rangle, \langle [1.5, 3.5), .25\rangle, \langle [1.5,4), .25\rangle\}.$$

$\mathcal{F}^E$ is consistent with

$$\mathbf{C}(\mathcal{F}^E) = [1.5, 2), \qquad \mathbf{U}(\mathcal{F}^E) = [1, 4].$$

The possibilistic histogram is the step function on the right of Fig. 4.2 with

$$\pi([1, 1.5)) = .5, \quad \pi([1.5, 2]) = 1, \quad \pi((2, 3.5]) = .5, \quad \pi((3.5, 4]) = .25$$

and $\pi(x) = 0$ elsewhere. The components of the focal set are

$$\vec{E} = \langle 1.5, 1, 1.5, 3.5, 2, 4\rangle, \qquad \vec{E}^l = \langle 1.5, 1, 1.5\rangle, \qquad \vec{E}^r = \langle 3.5, 2, 4\rangle,$$

$$\hat{E} = \langle 1, 1.5, 1.5, 2, 3.5, 4\rangle,$$

and the components of the possibilistic histogram are

$$E = \{1, 1.5, 2, 3.5, 4\}, \qquad E^l = \{1, 1.5\}, \qquad E^r = \{2, 3.5, 4\}.$$

with $Q = 5, Q^l = 2, Q^r = 3$ and the $G_k$ and $T_k$ identified in the figure. Since $\mathcal{F}^E$ is not consonant, the plausibility measure of $\mathcal{S}^E$ is not a natural possibility measure, but the constructed possibility measure $\Pi^*$ can be determined from (2.128).

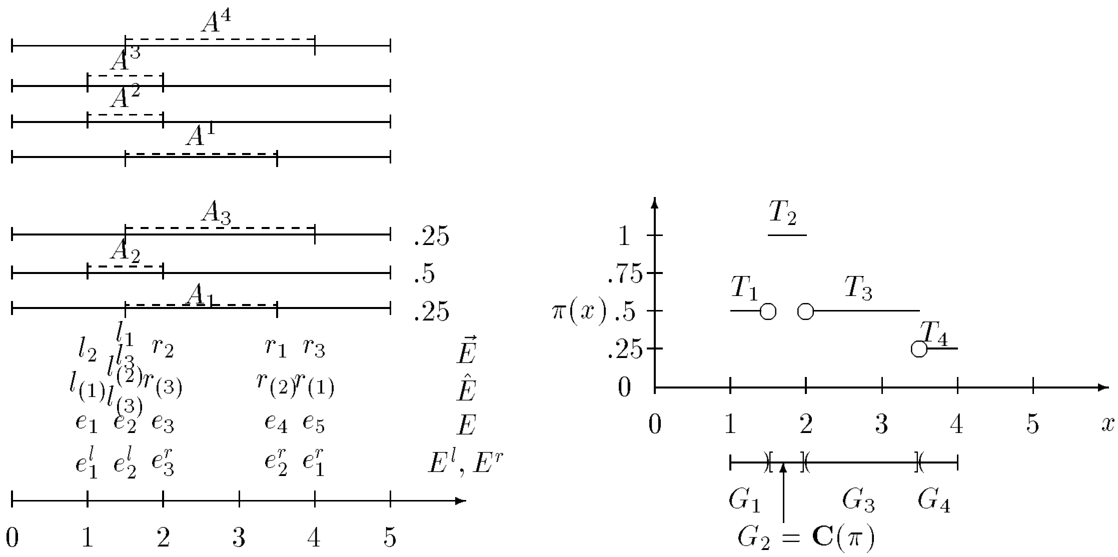


Figure 4.2: (Left) (top) A measurement record. (middle) $\mathcal{S}^E$ (bottom) Components of $\pi$. (Right) Possibilistic histogram $\pi$ with more components.

### 4.2.3 Possibilistic Histograms as Fuzzy Numbers

Possibilistic histograms are natural representations of possibility distributions. Since possibility theory is a weak representational form for uncertainty, it is appropriate that they produce meaningful forms of possibility distributions even given very few observations, as discussed in Sec. 3.3.3. In particular, possibilistic histograms are fuzzy intervals, and those with point cores are fuzzy numbers.

**Lemma 4.20** A possibilistic histogram $\pi$ is monotone nondecreasing from $-\infty$ to $\mathbf{C}(\pi)$ and monotone nonincreasing from $\mathbf{C}(\pi)$ to $\infty$.

**Proof:** Let $x \in \mathbb{R}$. The proof will be carried out for $x \in [-\infty, r_{(N)}]$. The remaining argument follows analogously for $x \in [l_{(N)}, \infty]$. Recall that endpoint ordering (4.16) carries over to the $e^l$ and $e^r$.

1. From (4.19–3), if $x < e^l_1$ then $\pi(x) = 0$.

2. Let $1 \leq k \leq Q^l$ and let $x_k \in G_k$, so that from (4.19–4) and the possibilistic histogram formula (4.12)

$$\pi(x_k) = \pi(e^l_k) = \sum_{A_j \ni x_k} \frac{C_j}{M} = \sum_{A_j \supseteq G_k} \frac{C_j}{M}.$$

From the consistent endpoint formula (4.15) and endpoint ordering (4.16),

$$\forall e^l_{k^l}, e^r_{k^r}, \quad e^l_{k^l} \leq e^l_{k^l+1} \leq e^l_{k^l+2} \leq e^r_{k^r}.$$

Therefore

$$|\{A_j : A_j \supseteq G_k\}| \quad \leq \quad |\{A_j : A_j \supseteq G_{k+1}\}|,$$

and so $\pi(x_k) \leq \pi(x_{k+1}) \leq 1$.

3. From (4.19–1), when $x \in [e^l_{Q^l}, e^r_{Q^r}] = \mathbf{C}(\pi)$, then $\pi(x) = 1$.

■

**Theorem 4.21** A possibilistic histogram $\pi$ is a fuzzy interval.

**Proof:** We need to show that both conditions of the definition of fuzzy interval (2.44) hold.

1. Possibilistic normalization follows from the definition of fuzzy set normalization (2.41), the sufficiency of consistency for possibilistic normalization (2.127), and the consistency of $\mathcal{F}^E$.

2. Convexity follows from the following three cases, which themselves follow from the lemma (4.20). Let $x, y, z \in \mathbb{R}, x \leq y \leq z$.

   (a) If $x \leq y \leq e^r_{Q^r}$ then $\pi(x) \wedge \pi(y) = \pi(x) \leq \pi(z)$.

   (b) If $e^l_{Q^l} \leq x \leq y$ then $\pi(x) \wedge \pi(y) = \pi(y) \leq \pi(z)$.

   (c) If $x \leq e^l_{Q^l} \leq e^r_{Q^r} \leq y$ then: if $x \leq z \leq e^r_{Q^r}$, then $\pi(x) \leq \pi(z)$; similarly, if $e^l_{Q^l} \leq z \leq y$, then $\pi(y) \leq \pi(z)$. Therefore $\pi(z) \geq \pi(x) \wedge \pi(y)$.

■

**Corollary 4.22** If $\exists x \in \mathbb{R}, \mathbf{C}(\mathcal{F}^E) = \{x\}$, then the possibilistic histogram $\pi$ is a fuzzy number.

**Proof:** Trivial from the definition of fuzzy number (2.45). ■

## 4.3 Continuous Approximations

Possibilistic histograms play the role in possibility theory that ordinary histograms do in traditional statistics. As maximum likelihood and other estimation methods are used in statistics to generate continuous approximations to histograms, so it is desirable to develop continuous or smooth approximations to possibilistic histograms.

**Definition 4.23 (Continuous Approximation)** Let $\bar{\pi}$ be a continuous possibility distribution which approximates a possibilistic histogram $\pi$.

One of the most significant differences between possibilistic and stochastic histograms is that the former are collections of the intervals $T_k$, not discrete points. Therefore, normal interpolation or approximation methods (such as curve-fitting or maximum-likelihood estimation) are not appropriate. Instead, a representative set of points from the intervals $T_k$ should be selected from $\pi$, and then a continuous curve $\bar{\pi}$ fitted to them.

### 4.3.1 Candidate Points

First it is necessary to characterize these candidate points from the possibilistic histogram.

**Definition 4.24 (Possibilistic Histogram Candidate Points)** Assume a possibilistic histogram considered as a locus of points

$$\pi := \{\langle e_k, \pi(e_k)\rangle\} \subseteq \mathrm{I\!R} \times [0,1].$$

Then denote:

- The left and right endpoints of each of the $T_k, 1 \leq k \leq Q-1$:

$$\mathbf{t}_k^l := \begin{cases} \langle e_k, \pi(e_k)\rangle, & e_k \in E^l \\ \langle e_k, \pi(e_{k+1})\rangle, & e_k \in E^r \end{cases}, \qquad \mathbf{t}_k^r := \begin{cases} \langle e_{k+1}, \pi(e_k)\rangle, & e_k \in E^l \\ \langle e_{k+1}, \pi(e_{k+1})\rangle, & e_k \in E^r \end{cases}.$$

- The midpoints of each of the $T_k, 1 \leq k \leq Q-1$:

$$\mathbf{h}_k := \left\langle \frac{e_k + e_{k+1}}{2}, \pi(e_k) \right\rangle.$$

- The midpoint of the core:

$$\mathbf{c} := \mathbf{h}_{Q^l} = \left\langle \frac{l_{(N)} + r_{(N)}}{2}, 1 \right\rangle.$$

- The endpoints of the support on the axis:

$$\mathbf{l} := \mathbf{t}_1^l = \left\langle l_{(1)}, 0 \right\rangle, \qquad \mathbf{r} := \mathbf{t}_{Q-1}^r = \left\langle r_{(1)}, 0 \right\rangle.$$

- The set of all the interval mid- and end-points to which a continuous curve *may* be fit:

$$\mathbf{K}' := \{\mathbf{t}_k^l, \mathbf{t}_k^r, \mathbf{h}_k\}.$$

- The set of all the interval mid- and end-points to which a continuous curve actually *will* be fit: $\mathbf{K} \subseteq \mathbf{K}'$.

- Finally, the set of *all* the points to which the curve will be fit:

$$\mathbf{D} := \{\mathbf{c}, \mathbf{l}, \mathbf{r}\} \cup \mathbf{K} \subseteq \pi.$$

The structure of $\mathbf{D}$ is then characterized by the following principle:

**Principle 4.25 (Candidate Point Selection)** $\mathbf{K}$ may be any subset of $\mathbf{K}'$ such that $\forall x \in \mathbf{U}(\pi)$, there is at most one point in $\mathbf{K}$ for which $x$ is the ordinate.

Note that $\mathbf{K} = \emptyset$ is allowed.

Both the definition (4.24) and the principle (4.25) are justified by the following argument:

1. Possibilistic normalization requires at least one point from the core to be a candidate. $\mathbf{c}$ is the only natural single point from the core, and so its requirement serves as the least restrictive normalization requirement.

2. For $\bar{\pi}$ to be zero outside the support $\mathbf{U}(\pi)$, and since $\bar{\pi}$ is continuous, $\bar{\pi}$ should drop to the axis through the points $\mathbf{l}$ and $\mathbf{r}$.

3. The above two criteria are the only *necessary* conditions to construct a continuous possibility distribution with support $\mathbf{U}(\pi)$. Therefore $\{\mathbf{c}, \mathbf{l}, \mathbf{r}\} \subseteq \mathbf{D}$, but $\mathbf{K}$ may be empty.

4. For each interval $T_k$, the naturally identifiable points, which are also consistent with the ordinal nature of possibilistic information, are $\mathbf{d}_k^l, \mathbf{t}_k^r$, and $\mathbf{h}_k$. Therefore they *may* be included in $\mathbf{K}$.

5. The final requirement in (4.25) is simply a statement that $\bar{\pi}$ must be a function, so that $\forall x \in \mathbf{U}(\bar{\pi}), \exists! \bar{\pi}(x)$. This would preclude, for example, including both the right limit of a $T_k$ open on the right and the left limit of $D_{k+1}$ closed on the left, which are equal in $x$ but differ in $\pi(x)$.

### 4.3.2 An Example

Consider the example in Fig. 4.3. The left side shows two intervals in dashed lines below the axis, each of which is observed once. The components of the $T_k$ with $N = M = 2, Q = 3$, and $\mathbf{c} = \mathbf{h}_2$ are also shown. $\mathbf{t}_1^l$ and $\mathbf{t}_3^r$ are excluded from $\mathbf{K}$ due to conflicts with $\mathbf{l}$ and $\mathbf{r}$, leaving a candidate set

$$\mathbf{K}' = \{\mathbf{h}_1, \mathbf{t}_1^r, \mathbf{t}_2^l, \mathbf{t}_2^r, \mathbf{t}_3^l, \mathbf{h}_3, \mathbf{t}_3^r\}.$$

Any subset $\mathbf{K} \subseteq \mathbf{K}'$ (including the empty set) can be chosen as long as it does not contain either set of conflicts $\{\mathbf{t}_1^r, \mathbf{t}_2^l\}$ or $\{\mathbf{d}_2^r, \mathbf{t}_3^l\}$.

### 4.3.3 Piecewise Linear Approximations

Once a set of points is selected, a variety of curve-fitting methods are available to determine $\bar{\pi}$. The simplest and most direct is to connect them with line segments, producing a piecewise linear, continuous distribution. Three of these are shown on the right of Fig. 4.3 for the sets

$$\mathbf{K} = \{\mathbf{h}_1, \mathbf{t}_2^l, \mathbf{t}_2^r, \mathbf{h}_3\}, \quad \emptyset, \quad \{\mathbf{t}_1^r, \mathbf{t}_3^l\},$$

moving from the outside to the inside. Alternatively, nonlinear regression or spline methods can be used to fit the selected points to one of the exponential or quadratic forms which are commonly used for fuzzy numbers [54, 168, 280].

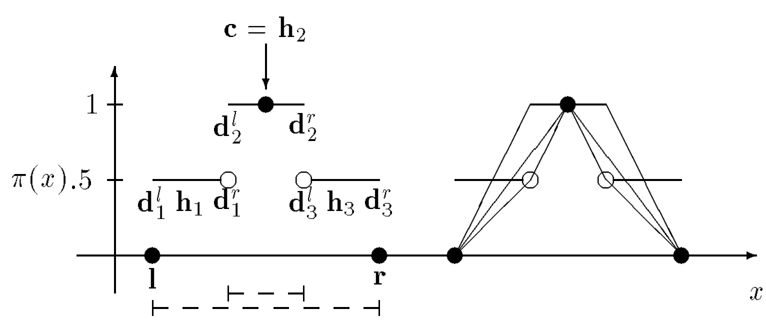


Figure 4.3: (Left) A simple possibilistic with its candidate points. (Right) Three example piecewise linear continuous approximations.

An advantage of the line-segment method is that even given very few observations, $\bar{\pi}$ has the same form as the fuzzy intervals and numbers typically used in fuzzy systems applications. Some of these are shown in Fig. 4.4, with some example observed intervals below them which could give rise to them. Case $A$ is a square distribution produced by a single crisp interval $[a, b]$; $B$ is the triangular form, produced in all cases when $d = \mathbf{c}$ and $\mathbf{K} = \emptyset$; $C$ is the outermost case of Fig. 4.3 for the observations $[f, i], [g, h]$.

In case $D$ it is also common for $\pi$ to extend to the right by letting $m \longrightarrow \infty$, so that $\forall x \geq l, \pi(x) = 1$. Either condition can result when point observations $j, k, l$ are interpreted either as distances from a fixed $m$ (perhaps an upper bound), or as magnitudes in relation to one or the other infinities. In this last case, $\pi$ is simply equivalent to a cumulative probability distribution; but this approach is in keeping with the ordinal possibilistic concepts of capacity, distance, and similarity.

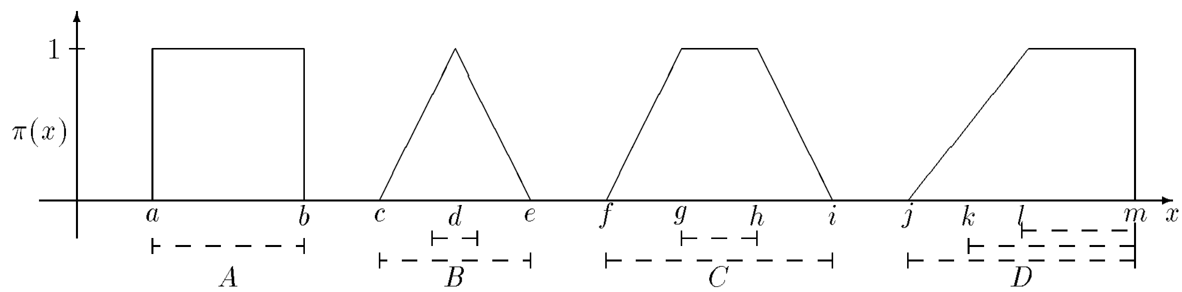


Figure 4.4: Typical fuzzy intervals and numbers used in applications.

## 4.4 Compatibility of Possibilistic Histograms

We have seen that it is natural and appropriate to derive possibility distributions, in the form of possibilistic histograms, from consistent or consistently transformed random sets. Since in Sec. 3.4.2.3 attempts to derive possibility distributions from probability distributions and specific random sets were criticized, it is important to consider how set-valued data are viewed from a traditional information-theoretic perspective. In particular, given a possibilistic histogram, compatibility with related probability distributions and the semantic criteria of Sec. 3.2 should be examined.

### 4.4.1 Possibilistic Histograms and the Possibility of Occurrence

The primary semantic criteria from Sec. 3.2.4 was that occurrence of an event requires maximal (unitary) possibility. In a possibilistic histogram the occurring events are exactly those $A^s \in \vec{A}$ which have been observed. So this condition is easily met by possibilistic histograms.

**Corollary 4.26** If $\mathcal{F}^E$ is consistent, then $\forall A^s \in \vec{A}, \Pi(A) = 1$.

**Proof:** Fix $A^s$. Then $C(A^s) \geq 1$, so $m(A^s) \geq 1/M > 0$. The result follows from the theorem (2.71) and the consistency of $\mathcal{F}^E$. ■

### 4.4.2 GK-Compatible Probability Distributions

Probability distributions which conform to PPC for a possibilistic histogram should also be considered. Under the PPC principle (3.5), it is necessary that $p(\omega) = 0$ wherever $\pi(\omega) < 1$, that is $\forall \omega \notin \mathbf{C}(\mathcal{F}^E)$. In the example in Fig. 4.2, that would yield $p > 0$ only on the interval $[1.5, 2)$. No further information would be provided by $\pi$, and so the MEP would yield the uniform probability density

$$p^*(\omega) = \begin{cases} 2, & \omega \in [1.5, 2) \\ 0, & \text{elsewhere} \end{cases} .$$

This result makes complete sense in the context of the nature of subset measurements. Given a consistent set of observed intervals, if they are all to be believed then all that can be said is that the event actually happened *somewhere* in the core. There the possibility is unitary, and by PPC the probability is positive. But there is no further information about the *likelihood* of the event being anywhere *particular* inside the core, thus requiring the maximally uninformative probability distribution.

The fact that $\forall \omega \in \mathbf{U}(\pi), \omega \notin \mathbf{C}(\pi), 0 < \pi < 1$ indicates that it is *somewhat* possible for *another* observation, perhaps at another time, to be found somewhere between the core and the edge of the support, but not *completely* possible, since nothing can be said to have been actually observed there yet. Thus the subset measurements give *no* likelihood information about the occurrence of an $\omega$ in this region, and by PPC $p = 0$ there.

If $\mathcal{S}^E$ is inconsistent, and thus a consistent approximation must be made, then for a focus $\omega_0 \in \Omega, \mathbf{C}(\mathcal{S}^E) = \{\omega_0\}$, and so $p$ will be a Dirac-delta function at $\omega_0$.

### 4.4.3 Frequency Distributions from Empirical Random Sets

It is also interesting to see how a purely "probabilistic" treatment would approach set-statistics. In particular, it is possible to use other counting methods to derive an ordinary frequency distribution $\vec{f}$ from the counts attached to each observed subset.

#### 4.4.3.1 Frequency Analysis of Subset Measurements

In order to simplify the problem, consider the case of two overlapping observations on a discrete universe. Let $\Omega = \{a, b, c\}$, and assume two observations $A^1 = \{a, b\}$ and $A^2 = \{b, c\}$, so that $C(A^1) = C(A^2) = 1$.

On a pure frequency analysis at the level of the subsets $A^s$, then $\Pr(A^1) = \Pr(A^2) = 1/2$. Under the assumption that Pr should have an additive probability distribution $p\colon \Omega \mapsto [0, 1]$, then

$$\begin{array}{ccccccl} p(a) & + & p(b) & + & p(c) & = 1 \\ p(a) & + & p(b) & & & = 1/2 \\ & & p(b) & + & p(c) & = 1/2 \end{array}$$

which has the solution $p(a) = p(c) = 1/2, p(b) = 0$. This is entirely unsatisfactory, and maximally incompatible with the possibilistic results above: it *eliminates* probability exactly on $b$, the point where there is the *most* evidence, and where in the possibilistic histogram $\pi(b) = 1$. Only slightly more complicated cases, such as the example in Sec. 4.2.2, reveal that this method frequently does not yield *any* feasible solutions for non-negative probabilities.

#### 4.4.3.2 Subset to Element Counts

Another approach is to translate the counts on subsets into counts on elements, thus establishing a mapping $C \mapsto c$. There are a number of ways in which that could be done.

**Duplicated Counts** We could say that a nonspecific observation is really an observation of *every* element of the subset. Then each observation of a subset $A^s$ would contribute one element count for every $\omega \in A^s$. Then the overall element count is

$$\forall \omega \in \Omega, \quad c(\omega) = \sum_{A_j \ni \omega} C_j. \tag{4.27}$$

**Corollary 4.28**

$$f(\omega) = \frac{c(\omega)}{\sum_{A_j \in \mathcal{F}^E} C_j |A_j|}.$$

**Proof:**

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega \in \Omega} c(\omega)} = \frac{c(\omega)}{\sum_{\omega \in \Omega} \sum_{A_j \ni \omega} C_j} = \frac{c(\omega)}{\sum_{A_j \in \mathcal{F}^E} C_j |A_j|}.$$

■

By this method, the example in Sec. 4.2.2 yields the frequency distribution

$$\vec{f} = \langle 2/9, 4/9, 2/9, 1/9 \rangle .$$

Note that this is identical to $\vec{\pi}$ for elements having the same numerator, but the denominator changed from 4 (which is $\sum C_j$) to 9 (which is $\sum c(\omega) = \sum C_j |A_j|$).

In fact, the effect of this count duplication method is to establish a maximum normalized ratio scale (Sec. 3.4.2.2) between $\pi$ and $f$.

**Theorem 4.29** Given a consistent $\mathcal{F}^E$ with a frequency distribution $f$ determined by (4.27), then $\forall \omega \in \Omega$,

$$f(\omega) = \frac{\pi(\omega)}{\sum \pi(\omega)}, \qquad \pi(\omega) = \frac{f(\omega)}{\bigvee f(\omega)}$$

**Proof:** From the possibilistic histogram formula (4.12) and (4.27),

$$\forall \omega \in \Omega, \quad M\pi(\omega) = \sum_{A_j \ni \omega} C_j = c(\omega).$$

Therefore from the corollary (4.28),

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega\in\Omega} c(\omega)} = \frac{M\pi(\omega)}{\sum_{\omega\in\Omega} M\pi(\omega)} = \frac{\pi(\omega)}{\sum_{\omega\in\Omega}\pi(\omega)}.$$

The second result follows from the converse of the ratio scale frequency conversion (3.17). ■

Thus the disadvantages of duplicating counts like this are clear. First, frequency additivity is violated because

$$\sum_{\omega\in A_j} c(\omega) = \sum_{\omega\in A_j}\sum_{A_k\ni w} C_k \geq C_j.$$

Also, GK-compatibility is generally violated in virtue of the ratio scale frequency conversion, as discussed in Sec. 3.4.2.2.

**Distributed Counts** Instead of a subset count contributing multiple element counts, the single subset count can be additively distributed amongst the $\omega \in A$. Since there is no further information about how to distribute the count, then by the MEP a uniform distribution should be used. Then the element count for each $\omega \in \Omega$ is

$$\forall\omega\in\Omega,\quad c(\omega) = \sum_{A_j\ni\omega}\frac{C_j}{|A_j|}. \tag{4.30}$$

**Corollary 4.31** $f(\omega) = c(\omega)/M$.

**Proof:** Because

$$\sum_{\omega\in\Omega} c(\omega) = \sum_{\omega\in\Omega}\sum_{A_j\ni\omega}\frac{C_j}{|A_j|} = \sum_{A_j\in\mathcal{F}^E}\frac{C_j|A_j|}{|A_j|} = \sum_{A_j\in\mathcal{F}^E} C_j = M,$$

therefore

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega\in\Omega} c(\omega)} = \frac{c(\omega)}{M}.$$

■

By this method, the example in Sec. 4.2.2 yields a frequency distribution

$$\vec{f} = \langle 1/4, 11/24, 5/24, 1/12\rangle.$$

Not surprisingly, this method is closely related to the applications of the MEP as discussed in Sec. 2.6.4.2, Sec. 2.8.3, and Sec. 3.4.2.2 (3.26).

**Theorem 4.32** Assume an empirical random set $\mathcal{S}^E$ and let $f$ be a frequency distribution determined by (4.30). Then $f$ is the maximum entropy probability distribution $p^{\mathcal{S}^E}$ from (2.122).

**Proof:** From (4.30), (4.31), the set-frequency definition (4.4), and the maximum entropy probability distribution formula (2.122), then $\forall \omega \in \Omega$,

$$f(\omega) = \frac{c(\omega)}{M} = \sum_{A_j \ni \omega} \frac{C_j}{|A_j| M} = \sum_{A_j \ni \omega} \frac{m_j^E}{|A_j|} = p^{\mathcal{S}^E}(\omega).$$

■

## 4.5 Sources of Set Statistics

While the previous sections have established the mathematical basis for possibilistic measurement, it remains to be seen what methodological and pragmatic grounds can generate subset counts, and thus empirical random sets. What are, in fact, the sources of set statistics?

The disjoint structure of the measurable classes of classical instruments naturally admits to data governed by stochastics. This is only because the disjointness of $\mathcal{C}$ allows the set-based frequency distribution $m^E : 2^\Omega \mapsto [0,1]$ to be regarded as a traditional frequency distribution $f'$ on a new domain $\Omega' = \mathcal{C}$. However, in the general case of a non-disjoint $\mathcal{C}$ this does not hold.

### 4.5.1 Instrument Ensembles

One way to generate measurements which are intersecting subsets is to use a *variety* of classical instruments.

**Definition 4.33 (Instrument Ensemble)** Let an **instrument ensemble** $\mathbf{F} := \{\mathcal{C}^s\}$ be a family of disjoint measuring devices such that

$$\mathcal{C}^s := \left\{ A^s_{j'^s} \right\}, \qquad 1 \le s \le M, \quad 1 \le j'^s \le N'^s := |\mathcal{C}^s|,$$

$$\forall A^s_{j_1'^s}, A^s_{j_2'^s} \in \mathcal{C}^s, \qquad A^s_{j_1'^s} \perp A^s_{j_2'^s}.$$

Denote $\mathcal{C}^{\mathbf{F}} := \bigcup_{s=1}^M \mathcal{C}^s$ as the general measuring device derived from $\mathbf{F}$.

An ensemble $\mathbf{F}$ can be considered as either multiple, heterogeneous instruments taking separate measurements at the same time, or as a single instrument which is changing its structure over time. While each of the $\mathcal{C}^s$ is disjoint, of course their combination in $\mathcal{C}^{\mathbf{F}}$ is not.

A natural partial order is defined on $\mathbf{F}$.

**Definition 4.34 (Ordering of Disjoint Instruments)** Given $\mathbf{F}$, let $\mathcal{C}^1 \preceq \mathcal{C}^2$, so that $\mathcal{C}^1$ refines $\mathcal{C}^2$ and $\mathcal{C}^2$ coarsens $\mathcal{C}^1$, when

$$\forall A^2_{j'2} \in \mathcal{C}^2, \quad \exists \left\{ A^1_{j'1} \right\} \subseteq \mathcal{C}^1, \quad A^2_{j'2} = \bigcup A^1_{j'1}.$$

As an example, $\mathcal{C}^1$ could be a thermometer reading in tenths of degrees, while $\mathcal{C}^2$ could be a mutually calibrated thermometer reading in whole degrees.

**Definition 4.35 (Consonant Ensemble)** $\mathbf{F}$ is **consonant** if all the $\mathcal{C}^s \in \mathbf{F}$ are comparable under $\preceq$.

If $\mathbf{F}$ is consonant, then without loss of generality for ordering, let $\mathcal{C}^1 \preceq \mathcal{C}^2 \preceq \cdots \preceq \mathcal{C}^M$. $\mathcal{C}^1$ is the most refined, the most accurate of the instruments.

#### 4.5.1.1 Random Sets from Instrument Ensembles

Random sets arise naturally from instrument ensembles when a single sample is taken from each instrument in the ensemble.

**Proposition 4.36** Assume an instrument ensemble $\mathbf{F}$ so that $\mathcal{C}^{\mathbf{F}}$ is a general measuring device, and let $A^s$ be a single subset observed in device $\mathcal{C}^s$. Then the vector of observations over $\mathbf{F}$ is $\vec{A}$, and $\mathcal{F}^E, m^E$, and $\mathcal{S}^E$ can be derived as in Sec. 4.1.3.

If any of the $\mathcal{C}^s$ share common members (in particular, if any of their elements are equal), then some of the $A^s$ may be equal, yielding multiple observations in $\vec{A}$ of certain subsets. Otherwise, all subsets will be observed a single time, but will not necessarily be disjoint.

**Theorem 4.37** If $\mathbf{F}$ is consonant and $\mathcal{F}^E$ consistent, then $\mathcal{F}^E$ is a nest.

**Proof:** Let $M = 2$ so that $\mathcal{C}^1 \preceq \mathcal{C}^2$. Let $A^1 \in \mathcal{C}^1$ and $A^M = A^2 \in \mathcal{C}^M = \mathcal{C}^2$. If $\mathcal{F}^E$ is consistent, then $A^1 \not\perp A^2$. But from the definition (4.34), $\exists \left\{ A^1_{j'1} \right\} \subseteq \mathcal{C}^1, A^2 = \bigcup A^1_{j'1}$. Therefore $A^1 \subseteq A^2$. By induction to $2 < M \in \mathcal{W}$, then $\forall A^s \in \vec{A}, A^1 \subseteq A^2 \subseteq \cdots \subseteq A^M$. Since $\forall A_j \in \mathcal{F}, \exists A^s \in \vec{A}$, therefore $\forall A_j \in \mathcal{F}^E, A_1 \subseteq A_2 \subseteq \cdots \subseteq A_N$. ∎

Of course, in this case a possibilistic analysis is less useful than it would be in others, since there is an absolute gain in accuracy in the move towards the finest measurements of $\mathcal{C}^1$. Nevertheless, the mathematical analysis is available.

#### 4.5.1.2 An Example

Fig. 4.5 shows an example of an instrument ensemble which can result in the measurement record and possibilistic histogram in the example in Sec. 4.2.2. Let $\Omega = [0, 5] \subseteq \mathbb{R}$ and define a family $\mathbf{F}$ of four measuring devices

$$\mathcal{C}^1 = \{[0,1),[1,2),[2,3),[3,4),[4,5]\}, \qquad \mathcal{C}^2 = \{[0,1),[1,2),[2,3.5),[3.5,5]\},$$
$$\mathcal{C}^3 = \{[0,1.5),[1.5,3.5),[3.5,4),[4,5]\}, \qquad \mathcal{C}^4 = \{[0,1.5),[1.5,4),[4,5]\},$$

so that $M = 4$ and

$$\mathcal{C}^{\mathbf{F}} = \{[0,1),[0,1.5),[1,2),[1.5,3.5),[1.5,4),[2,3),[2,3.5),[3,4),[3.5,4),[3.5,5],[4,5]\}.$$

$\mathbf{F}$ is not consonant, but $\mathcal{C}^3 \preceq \mathcal{C}^4$. The intervals shown would result in the measurement record from Sec. 4.2.2.

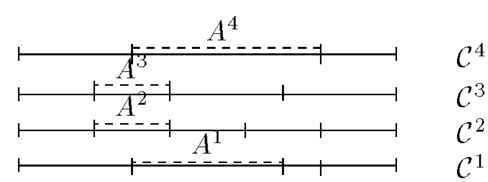


Figure 4.5: An example instrument ensemble.

#### 4.5.1.3 Consequences of Instrument Ensembles

Consider two observations from different devices $A^1 \in \mathcal{C}^1$ and $A^2 \in \mathcal{C}^2$. It can be expected that $A^1 \not\perp A^2$. In the event that $A^1 \perp A^2$, then there can be no value in agreement between the two instruments, so that at least one of the devices $\mathcal{C}^1$ or $\mathcal{C}^2$ would be regarded as being in error, or perhaps even the assumption of the "reality" of the quantity being measured would be questioned.

So while there is nothing in the mathematics that would preclude such a result, pragmatic conditions require that $\mathcal{F}^E$ be consistent, so that $\mathcal{S}^E$ has a natural possibility distribution $\pi$ and at worst the constructed possibility measure $\Pi^*$ from (2.128). In the event that $\mathcal{F}^E$ is nevertheless not consistent, and there are pragmatic reasons for accepting the results of the measurement, then the possibilistic normalization methods outlined in Sec. 2.8 are available to construct consistent random sets.

Gathering statistics on an ensemble of instruments $\mathbf{F}$ is a fundamentally different process from gathering time-series data on a single classical instrument $\mathcal{C}$. The former is governed by random set criteria, the latter by stochastic criteria. In particular, the backdrop of the measurements is changed from multiple time trials on a single instrument $\mathcal{C}$ to multiple trials over a population of *instruments* $\mathcal{C}^s$ at a *single* time.

From the comments in Sec. 3.3.6, we can see the appropriateness of instrument ensembles for possibilistic measurement on complex systems. If a measurement on a complex system irreversibly perturbs it, then it may be necessary to change the measuring device to accommodate the new form of the system, thus creating time-data from a changing instrument. If getting multiple time samples is difficult, then using multiple heterogeneous instruments may allow the extraction of more information from measurements at a single time.

The instrument ensemble approach described here is very similar to the "hidden labels" method described by Lemmer [176].

#### 4.5.1.4 Sources of Instrument Ensembles

It is necessary to consider some possible sources of instrument ensembles.

**Mutually Discalibrated Instruments:** An instrument ensemble can result when multiple instruments are used to measure a single quantity, but they are all differently calibrated, either in precision, phase, or both. Continuing with the thermometer examples used above, one may be calibrated in degrees from 0 to 100, and another in third degrees from 0.2 to 100.2.

**Multiple Traditional Uncertainty Intervals:** Traditional methods for representing measurements involve uncertainty intervals around point measurements. For example, an uncertainty interval might be expressed as a $3\sigma$ standard deviation or a 95% confidence interval around a point [8, 282]. Thus measurement results expressed in this form are naturally subset observations. So when multiple such measurements are made, as, for example, when multiple teams measure the same fundamental physical constant [209], then an empirical random set naturally results.

**Indirect Measurement:** Sometimes some quantity $a$ cannot be measured directly, but rather knowledge about it must be gained by inference from the knowledge of other quantities $b$ and $c$, which can be measured. Even if $b$ and $c$ are measured using classical (disjoint) instruments, still each results in an interval in the range of $b$ or $c$. So when inference is made from each of these measurements back to the state of $a$, then each will result in turn in an interval in the range of $a$, creating an empirical random set. Hopefully this set will be consistent, as argued above indicating the "reality" of the attribute $a$ and/or the validity of the inference process.

### 4.5.2 Point Data from Specific Instruments

As discussed in Sec. 4.1.4, any classical (disjoint) instrument generates observations from a disjoint class of intervals $\mathcal{C} = \{A_{j'}\}, A_{j'} \subseteq \Omega \subseteq \mathbb{R}$ which can thereby be regarded as distinct points $A_{j'}$ in a higher-level state space $\Omega' = \mathcal{C}$.

In the sequel of this section we will change notation somewhat, and consider a single measuring device which yields observations as actual data points in a lower-level state space, a closed interval $\Omega \subseteq \mathbb{R}$. We will try to follow the notation as used in Sec. 4.1 as much as possible, especially concerning indices and bounds.

**Definition 4.38 (Datum, Data Stream, Data Set)**

- Denote an observation as a **datum** $d \in \Omega$.
- The **data stream** is the vector

$$\vec{D}' = \left\langle d^t \right\rangle, \qquad 1 \leq t \leq M' := |\vec{D}'|.$$

- Not all of the methods presented below will use the entire data stream. A subcollection of data is the **reduced data stream**, a sub-vector of $\vec{D}$

$$\vec{D} = \langle d^s \rangle, \qquad 1 \leq s \leq M := |\vec{D}| \leq M'$$

  where $\forall d^s \in \vec{D}, d^s \in \vec{D}'$.

- The set generated by eliminating duplicates in $\vec{D}$ is the **data set**

$$D = \{d_j\}, \qquad 1 \leq j \leq N := |D| \leq M,$$

  such that

$$\forall d_j \in D, \quad \exists d^s \in \vec{D}, \quad d_j = d^s.$$

### 4.5.3 Order Statistical Methods

In Sec. 4.5.1 an empirical random set $\mathcal{S}^E$ was derived from set-statistics collected on an instrument ensemble (either a collection of heterogeneous instruments at a single time or a changing instrument over multiple times), and these non-specific data were contrasted with specific time-series data collected on a single, unchanging instrument. But even given a single measuring device and time-series data gathered on it, nonspecific data can still be generated by constructing intervals from the data points.

#### 4.5.3.1 Order Statistics

In approaching a possibilistic analysis of data, we must use semantic criteria which are natural and appropriate for possibility theory, not for probability theory. As discussed in Sec. 3.3.1.3, possibility is fundamentally an *ordinal* measure, relating to concepts of distance, similarity, intensity, and capacity; while probability relates to concepts of proportion, likelihood, and frequency.

A possibilistic analysis of $\vec{D}$ will be approached through the **order statistics** of $\vec{D}$ [44].

**Definition 4.39 (Order Statistics)** Given a reduced data stream $\vec{D}$, the **order statistics** $d^{(s)}$ are a permutation of the $d^s \in \vec{D}$ such that

$$d^{(1)} \leq d^{(2)} \leq \cdots \leq d^{(M)}.$$

$d^{(1)}$ and $d^{(M)}$ are the **extremes**, and denote the **range** interval

$$W := \left[d^{(1)}, d^{(M)}\right] \subseteq \Omega.$$

Also denote the order statistics of the data set $D$ as $d_{(j)}$ so that

$$d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(N)}.$$

**Definition 4.40 (Disjoint Intervals)** Given a data set $D$, denote the **disjoint intervals**

$$\delta_j := \left[d_{(j)}, d_{(j+1)}\right), \qquad 1 \leq j \leq N-1.$$

For completeness denote

$$\delta_N := \left[d_{(N)}, d_{(N)}\right] = \left\{d_{(N)}\right\}.$$

Finally let the set of disjoint intervals be $\Delta := \{\delta_j\}, 1 \leq j \leq N$.

**Corollary 4.41** $\Delta$ is a disjoint measuring device, and the $\delta_j$ partition $W$, so that $\bigcup_{j=1}^{N} \delta_j = W$.

**Proof:** Obvious from the definition of a disjoint device (4.7) and the disjoint intervals (4.40). ∎

#### 4.5.3.2 Focused Data Intervals

It is clear that $\Delta$ cannot provide a possibility distribution because it has no focus, nor any core. Therefore this method is dependent on the predication of a focus $\omega^* \in \Omega$, here denoted $u \in W$. The purpose of $u$ is to provide a value on which all the intervals (yet to be determined) agree; a value for which $\pi(u) = 1$. In this section the entire data set will be used, so let $\vec{D} := \vec{D}'$ and $M := M'$, and the single index $s$ will be used.

In keeping with possibilistic concepts, we are interested in deriving a possibility distribution based on the intervals between the various $d^{(s)}$ and the focus $u$, which naturally divides $W$ into left and right sub-intervals.

**Definition 4.42 (Focused Data Intervals)** Given a data stream $\vec{D}$ and focus $u \in W$, define the following:

- Left and right **range intervals**:

$$W^l := \left[d^{(1)}, u\right) = \left[d_{(1)}, u\right), \qquad W^r := \left(u, d^{(M)}\right] = \left(u, d_{(N)}\right].$$

- **Focused data intervals** for $1 \le s \le M$ and $1 \le j \le N$:

$$A^s := \begin{cases} \left[d^{(s)}, u\right], & d^{(s)} \in W^l \\ \left[u, d^{(s)}\right], & d^{(s)} \in W^r \\ [u, u], & d^{(s)} = u \end{cases}, \qquad A_j := \begin{cases} \left[d_{(j)}, u\right], & d_{(j)} \in W^l \\ \left[u, d_{(j)}\right], & d_{(j)} \in W^r \\ [u, u], & d_{(j)} = u \end{cases}.$$

- **Left and right focal sets**:

$$\mathcal{F}^l := \{A_j : d_{(j)} \in W^l\}, \qquad \mathcal{F}^r := \{A_j : d_{(j)} \in W^r\}.$$

- **Empirical focal set**: $\mathcal{F}^E := \mathcal{F}^l \cup \mathcal{F}^r$.

- For **subset counts**, let $C(A_j)$ be the number of occurrences of $d_{(j)}$ in $\vec{D}$, which is also the number of $A^s$ equal to $A_j$.

- **Empirical random set**: $\mathcal{S}^E$ determined from $\mathcal{F}^E$ and $C$ as in (4.6).

**Corollary 4.43** $\mathcal{F}^l$ and $\mathcal{F}^r$ are nests, $\mathcal{F}$ is a consistent focal set, and $\mathcal{S}^E$ is a consistent random set.

**Proof:** It is obvious that

$$d^{(s_1)}, d^{(s_2)} \in W^l, \quad s_1 \le s_2 \quad \rightarrow \quad A^{s_2} \subseteq A^{s_1},$$

$$d^{(s_1)}, d^{(s_2)} \in W^r, \quad s_1 \leq s_2 \quad \to \quad A^{s_1} \subseteq A^{s_2},$$

so that the first result follows from the definition of nest (2.88). Then from the definition of focused data intervals (4.42), $\forall 1 \leq s \leq M, u \in A^s$, and because $\forall d^{(s)}, d^{(s)} \in W^l$ or $d^{(s)} \in W^r$, therefore

$$\mathcal{F}^E = \mathcal{F}^l \cup \mathcal{F}^r = \{A_j\}$$

is consistent from the definition (2.66). The final result is trivial. ∎

Generally, each ordered datum $d^{(s)}$ will generate a single count for the interval $A^s$. However, if

$$\exists s_1, s_2, \quad d^{(s_1)} = d^{(s_2)}, \quad A^{s_1} = A^{s_2}$$

then multiple counts will be generated as discussed in Sec. 4.1.3.

**Corollary 4.44** If $u = d^{(1)}$ or $u = d^{(M)}$ then $\mathcal{S}^E$ is consonant.

**Proof:** If $u = d^{(1)}$, then from the definition (4.42) $W^l = [d^{(1)}, u) = [u, u) = \emptyset$, and so $\mathcal{F}^l = \emptyset$, so that $\mathcal{F}^E = \mathcal{F}^r$, which from (4.43) is a nest, so that $\mathcal{S}^E$ is consonant. ∎

**An Example** Consider the example in Fig. 4.6. Let $\Omega = [0, 5]$, and assume that six point observations in $\Omega$ are taken giving the data stream $\vec{D} = \langle 2, 1, 4, 1.5, 2, 4.5 \rangle$. The order statistics and ranges are

$$d^{(1)} = 1, \quad d^{(2)} = 1.5, \quad d^{(3)} = d^{(4)} = 2, \quad d^{(5)} = 4, \quad d^{(6)} = 4.5,$$

$$W^l = [1, u], \qquad W^r = [u, 4.5], \qquad W = [1, 4.5].$$

The corresponding data set is $D = \{1, 1.5, 2, 4, 4.5\}$ so that $N = 5 < M = 6$, with order statistics and disjoint intervals

$$d_{(1)} = 1, \quad d_{(2)} = 1.5, \quad d_{(3)} = 2, \quad d_{(4)} = 4, \quad d_{(5)} = 4.5$$

$$\Delta = \{[1, 1.5), [1.5, 2), [2, 4), [4, 4.5), [4.5, 4.5]\}$$

Assume that $u \in [2, 4]$, then the focused data intervals $A^s$ and $A_j$ are respectively

$$[1, u], [1.5, u], [2, u], [2, u], [u, 4], [u, 4.5]$$

$$[1, u], [1.5, u], [2, u], [u, 4], [u, 4.5]$$

so that the focal and random sets are

$$\mathcal{F}^E = \mathcal{F}^l \cup \mathcal{F}^r = \{[1, u], [1.5, u], [2, u]\} \cup \{[u, 4], [u, 4.5]\},$$

$$\mathcal{S}^E = \{\langle [1, u], 1/6 \rangle, \langle [1.5, u], 1/6 \rangle, \langle [2, u], 1/3 \rangle, \langle [u, 4], 1/6 \rangle, \langle [u, 4.5], 1/6 \rangle\},$$

This is illustrated on the left of Fig. 4.6, for $u = 3$, and the resulting possibility distribution is shown on the right.

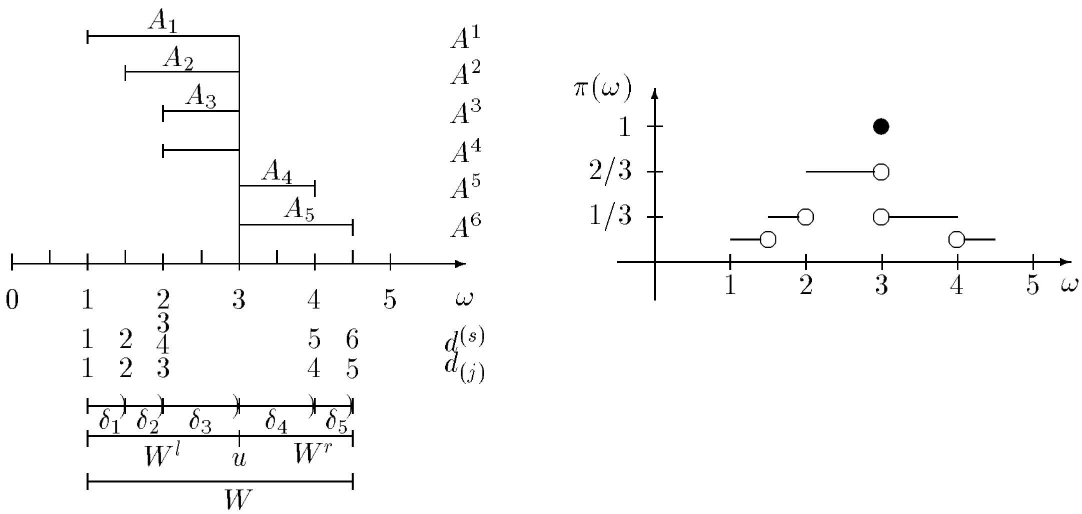


Figure 4.6: (Left) Consistent family from focused data set. (Right) Resulting possibilistic histogram.

**Choice of Focus** So far the method by which the focus $u$ can be chosen has not been discussed. While a number of methods suggest themselves, further selection of methods will depend on user methodology and further empirical research. However, note that the first four methods mentioned all yield $u \in [2, 4]$ in our example, which is the inner interval of $\Delta$ (see Sec. 4.5.3.3).

**Sample Mean** Selection of

$$u = \bar{D} = \frac{\sum d_s}{M}$$

is a possibility, although one which is not in keeping with possibilistic concepts. In the example, this would yield $u = 2.5$.

**Range Midpoint** The midpoint of $W$, denoted $\bar{W}$, is much more in keeping with possibilistic concepts:

$$u = \bar{W} := \frac{d^{(1)} + d^{(M)}}{2}.$$

This expresses something like the concept of a "possibilistic sample mean", and yields $u = 2.75$ in the example.

**Closest to Range Midpoint** There may be some value in having $u$ actually be one of the data points, so that $u \in D$. This can be done by selecting that $d_j \in D$ closest to $\bar{W}$:

$$u = \min_{d_j \in D} |d_j - \bar{W}|.$$

This yields $u = 2$ in the example.

**Data-Set Midpoint** The middle point of the data set itself can be chosen:

$$u = \begin{cases} d_{(\frac{N+1}{2})}, & N \text{ odd} \\ d_{(\frac{N}{2})} \text{ or } u = d_{(\frac{N}{2}+1)}, & N \text{ even} \end{cases}.$$

yielding $u = 2$ in the example. Alternatively, if $N$ is even then the midpoint of the central interval can be selected:

$$u = \frac{d_{(\frac{N}{2})} + d_{(\frac{N}{2}+1)}}{2}.$$

**Information Principles** Finally, the Uncertainty Principles of Sec. 2.6.4 can be applied.

Selection of $u$ can be regarded as a problem of ampliative reasoning, of making an inductive inference beyond the given information. Then the MUP of Sec. 2.6.4.2 can be invoked, and thus $u$ would be chosen for which $\mathbf{T}(\mathcal{S}^E)$ or $\mathbf{T}(\pi)$ is maximal.

Alternatively, this entire focused intervals method can be regarded as a frequency transformation problem from the frequencies of the $d^s \in \vec{D}$ to a possibility distribution $\vec{\pi}$. Then the UIP or Minimal Information Distortion principle (2.146) can be invoked, which will state that $u$ should be chosen so as to make the total uncertainty $\mathbf{T}(\mathcal{S}^E)$ as close as possible to the entropy of the frequency distribution derived from $\vec{D}$.

While frequency conversion methods were criticized in Sec. 3.4.2.2, we are assuming here that only disjoint time-series data in the form $\vec{D}$ are available. Thus in this case frequency conversion methods are naturally justified.

#### 4.5.3.3 Interval Cores

A potential disadvantage of the focused interval methods in Sec. 4.5.3.2 is the reliance on a singleton-valued core set $\mathbf{C}(\mathcal{F}^E) = [u, u] = \{u\}$, while the other elements of the method are the intervals $A^s, \delta_j$ and $A_j$. Instead, methods which assume an interval-valued core can be considered. A disadvantage of these methods is that they may eliminate some data points, thus loosing some information from the resulting $\mathcal{S}^E$.

**Definition 4.45 (Interval Cores)**

- Assume a core

$$\mathbf{C} := \mathbf{C}(\pi) = [\mathbf{C}^l, \mathbf{C}^r] \subseteq W.$$

- Given a data stream $\vec{D}'$, let the reduced data stream be

$$\vec{D} = \vec{D}' - \{d^t \in \mathbf{C}\} + \{\mathbf{C}_l, \mathbf{C}_r\},$$

(where the operation $-/+$ of a set from a vector is just the elimination or concatenation of the appropriate points from the vector), so that

$$M = |\vec{D}| = M' - |\{d^t \in \mathbf{C}\}| + 2.$$

- The left and right range intervals are redefined as

$$W^l := \left[d_{(1)}, \mathbf{C}^l\right), \qquad W^r := \left(\mathbf{C}^r, d_{(N)}\right],$$

but the overall range

$$W := [d^{(1)}, d^{(M)}] = [d_{(1)}, d_{(N)}] = W^l \cup \mathbf{C} \cup W^r$$

is unchanged.

- Also redefine the intervals $A^s$ and $A_j$ as follows:

$$A^s := \begin{cases} \left[d^{(s)}, \mathbf{C}^r\right], & d^{(s)} \in W^l \\ \left[\mathbf{C}^l, d^{(s)}\right], & d^{(s)} \in W^r \end{cases}, \qquad A_j := \begin{cases} \left[d_{(j)}, \mathbf{C}^r\right], & d_{(j)} \in W^l \\ \left[\mathbf{C}^l, d_{(j)}\right], & d_{(j)} \in W^r \end{cases}.$$

- $\mathcal{F}^l, \mathcal{F}^r, \mathcal{F}^E$, and $\mathcal{S}^E$ are defined as in Sec. 4.5.3.2.

**Corollary 4.46** $\mathcal{F}^l$ and $\mathcal{F}^r$ are nests, and $\mathcal{F}^E = \mathcal{F}^l \cup \mathcal{F}^r$ is consistent with core

$$\mathbf{C}(\mathcal{F}^E) = \bigcap_{j=1}^{N} A_j = \mathbf{C}.$$

**Proof:** Follows from argument almost identical to the focused data interval result (4.43), replacing $u$ with $\mathbf{C}$, and noting that from the definition (4.45) that $\forall A^s, \mathbf{C} \subseteq A^s$. ∎

**An Example** The focused data interval example from (4.5.3.2) is modified as shown in Fig. 4.7 for a core $\mathbf{C} = [2, 4]$ (other cores are possible). Since $\{d^{(3)}, d^{(4)}, d^{(5)}\} \subseteq \mathbf{C}$, therefore $\vec{D} = \langle 1, 2.5, 4.5 \rangle$ and $M = N = 5$. The intervals $\Delta, A^s, A_j, W^l$, and $W^r$ are shown on the left of the figure, and the resulting possibilistic histogram on the right.

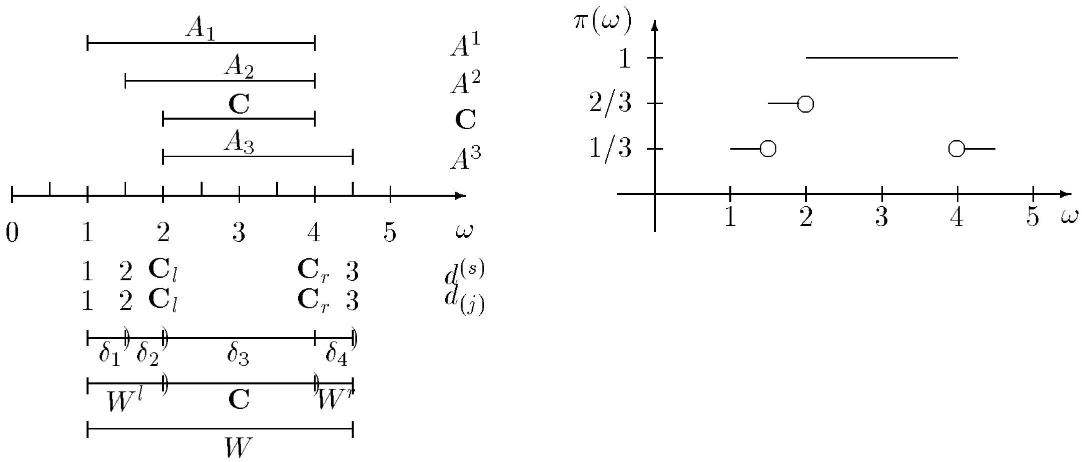


Figure 4.7: (Left) Reduced data set and stream from interval core. (Right) Resulting possibilistic histogram.

**Choice of Core** As with the selection of point focuses, there are a variety of methods by which an interval core can be selected.

**Central Disjoint Interval** If $N$ is even, then a central disjoint interval is naturally generated from the data set $D$:

$$\mathbf{C} = \delta_{N/2}.$$

Note that since $d_{(N/2)}, d_{(\frac{N}{2}+1)} \in \mathbf{C}$, all instances of them are eliminated from $\vec{D}'$ in forming $\vec{D}$.

**Modified Central Interval** If $N$ is odd, then there are two disjoint intervals on either side of $d_{(\frac{N+1}{2})}$. Thus we would select a core:

$$\mathbf{C} = \delta_{\frac{N-1}{2}} \cup \delta_{\frac{N+1}{2}}.$$

In the example this yields $\mathbf{C} = [1.5, 4]$. This method will eliminate instances of the three data points $d_{(\frac{N-1}{2})}, d_{(\frac{N+1}{2})}$, and $d_{(\frac{N+3}{2})}$ from $\vec{D}'$.

Alternatively, the midpoints of the two disjoint intervals around $d_{(\frac{N+1}{2})}$ can be selected as the endpoints of $\mathbf{C}$:

$$\mathbf{C} = \left[ \frac{d_{(\frac{N-1}{2})} + d_{(\frac{N+1}{2})}}{2}, \frac{d_{(\frac{N+1}{2})} + d_{(\frac{N+3}{2})}}{2} \right],$$

yielding $\mathbf{C} = [1.75, 3]$ in the example.

**Disjoint Interval Around a Focus** Given a method from Sec. 4.5.3.2 to select a point focus $u$, then $\mathbf{C}$ can just be selected as the data-generated disjoint interval around $u$:

$$\mathbf{C} = \delta_j, \quad u \in \delta_j.$$

As above, instances of $d_{(j)}$ and $d_{(j+1)}$ will be eliminated from $\vec{D}'$.

**Confidence Interval Around a Focus** It may be appropriate for the user to involve some traditional statistical information. Again, given some focus $u$, then $\mathbf{C}$ can be selected as the interval within a standard deviation of $u$:

$$\mathbf{C} = \left[ u - \sigma(\vec{D}), u + \sigma(\vec{D}) \right].$$

**Information Principles** Methods of Uncertainty Maximization or Invariance can be applied, as discussed in Sec. 4.5.3.2.

#### 4.5.3.4 Consonant Intervals from Focused Time-Series Data

Another disadvantage of the methods in Secs. 4.5.3.2 and 4.5.3.3 is that they yield consistent, but not consonant, families. Thus as discussed in Sec. 2.7.2, while they have natural maximum normalized possibility distributions $\pi$, nevertheless their plausibility measures Pl are not maximal, and thus are not possibility measures. The possibility measures $\Pi^*$ constructed from the distributions $\pi$ are not equal to the plausibilities.

Therefore it may be desirable to generate consonant families from a data stream $\vec{D}'$. However, as we move through the above methods (from consistent families with point focuses, through consistent families with interval cores, to consonant classes) the constraint on $\mathcal{S}^E$ increases, and more points are taken from the raw data stream $\vec{D}'$ to produce the reduced data stream $\vec{D}$, thus loosing information available in the original $\vec{D}'$. In point focus methods no points are lost, and roughly half the number of available intervals are lost in the following consonant methods. Thus as with the case of an ensemble of measuring devices discussed in Sec. 4.1.3, use of strictly consonant cases may be less useful than simply consistent cases.

Again, a number of methods present themselves.

**Inner Nested Intervals from Interval Core** Assume that an interval core $\mathbf{C} = [\mathbf{C}^l, \mathbf{C}^r]$ has been determined according to some method discussed in Sec. 4.5.3.3. Denote $A^1 = \mathbf{C}$, and denote a set of intervals $A^s = [A^s_l, A^s_r]$ such that $A^s_l, A^s_r \in D$ and $A^s \subseteq A^{s+1}$. Given an interval $A^s$, then $A^{s+1}$ is the nearest interval determined by $D$ containing $A^s$:

$$A^{s+1}_l = \max_{d_{(j)} \in D} d_{(j)} < A^s_l, \qquad A^{s+1}_r = \min_{d_{(j)} \in D} d_{(j)} > A^s_r.$$

The $A^s$ are available up to a maximal $A^N = W$. Together the focal set $\mathcal{F}^E = \{A^s\}$ form a consonant class. The count of $A^s$ can be determined as the maximum number of occurrences of either endpoint of $A^s$ in $\vec{D}$.

**Inner Nested Intervals from Point Focus** Assume instead that a point core $u \in W$ has been determined according to some method discussed in Sec. 4.5.3.2. Now simply let $A^1 = [u, u]$ and apply the method above.

**Outer Nested Intervals** Proceed in the opposite direction from above. Now define $A^1 = W$, and construct $A^{s+1}$ from $A^s$ as follows:

$$A_l^{s+1} = \min_{d_{(j)} \in D} d_{(j)} > A_l^s, \qquad A_r^{s+1} = \max_{d_{(j)} \in D} d_{(j)} < A_r^s.$$

### 4.5.4 Local Extrema

In order statistical methods any natural order of the data set in terms of the observation times or population against which the data are measured are deliberately ignored. Instead, the data stream $\vec{D}' = \langle d^t \rangle$ can be ordered strictly in terms of $t$, represented as a relation

$$\vec{D}' := \left\{ \left\langle t, d^t \right\rangle \right\} \subseteq \mathcal{W} \times \Omega,$$

and plotted as a time-series function of $t$. Then the data stream should be examined as to its *possibilistic* information, that is information related to the *breadth* of the data set, and the *extreme* values present.

Clearly if all we are interested in are the *global* extreme values of $\vec{D}'$ then

$$\max d^t \in \vec{D}', \qquad \min d^t \in \vec{D}'$$

alone could be examined, and so possibility 1 assigned to any $\omega$ between them. This is a much too coarse approach. Instead whatever series of strictly *local* extrema exist in $\vec{D}'$ can be used to generate interval observations.

**Definition 4.47 (Local Extrema)**

- A point $d^t \in \vec{D}'$ is a **local extremum** if any of the following conditions hold:

$$t = 1, \qquad t = M', \qquad d^{t-1} < d^t \text{ and } d^{t+1} < d^t, \qquad d^{t-1} > d^t \text{ and } d^{t+1} > d^t.$$

- Let the reduced data stream $\vec{D}$ be the vector of all local extrema, so that each $d^s$ is a local extremum.

- Let the observed intervals be those intervals between adjacent extrema:

$$A^s := \begin{cases} [d^s, d^{s+1}], & d^s < d^{s+1} \\ [d^{s+1}, d^s], & d^s > d^{s+1} \end{cases} .$$

Note that the local extrema must alternate, so that $\forall j$, if $e_j < e_{j+1}$ then $e_{j+1} > e_{j+2}$, and vice versa. Note also that each pair $A_j, A_{j+1}$ share exactly one endpoint.

#### 4.5.4.1 An Example

An example is provided for the data stream

$$\vec{D}' = \langle 1, 2, 4, 3, 5, 2.5, 1.5 \rangle$$

so that

$$M = 4, \qquad \vec{D} = 1, 4, 3, 5, 1.5, \qquad \mathcal{F}^E = \{[1,4], [3,4], [3,5], [1.5,5]\}.$$

as shown on the left of Fig. 4.8. The resulting possibility distribution is also shown in the right of the figure.

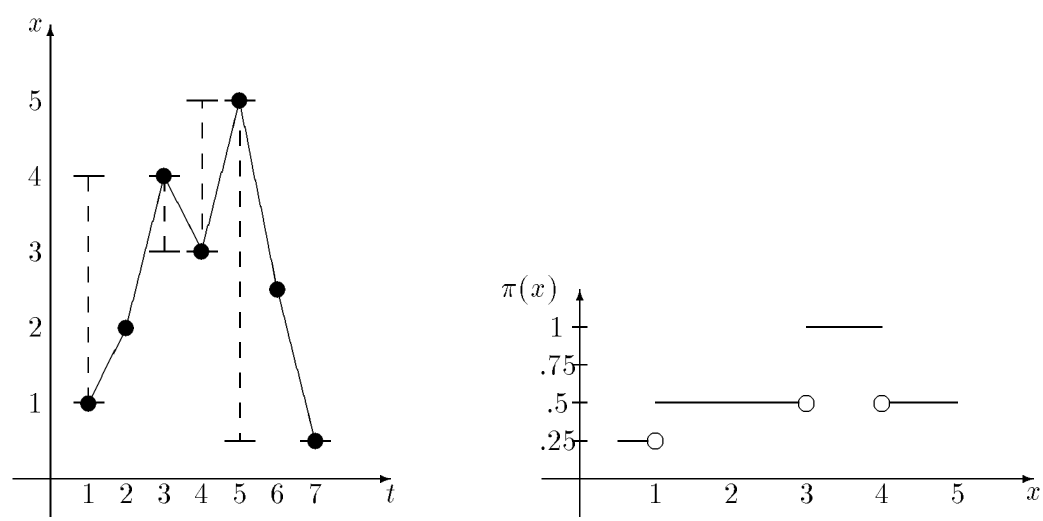


Figure 4.8: Observed intervals and resulting possibilistic histogram from local extrema.

#### 4.5.4.2 Properties

Note how much the possibilistic histograms generated according to this method reflect the properties of possibilistic mathematics, and not those of traditional, additive frequency distributions. In the example shown in Fig. 4.9, the single point for $t = 7$ affects the distribution as much as the other six points together.

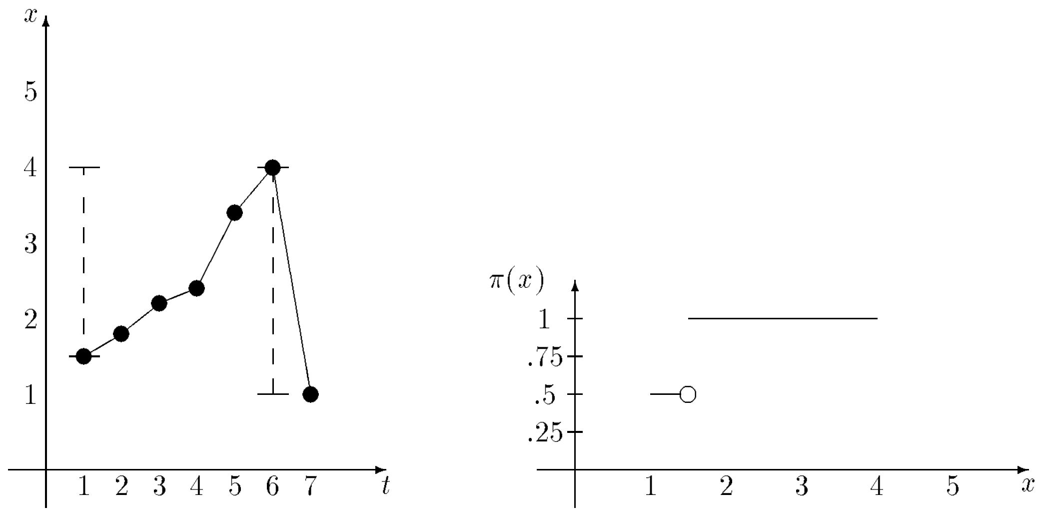


Figure 4.9: An extreme example showing possibilistic properties.

Nor is there any guarantee that the random sets produced by this method will be consistent. Possibilistic normalization may be required.

# Chapter 5

# Possibilistic Processes

Other things being equal, a monotonous
environment facilitates mechanisation.

— *Arthur Koeslter*

Possibilistic models were characterized in Sec. 3.1 and especially Sec. 3.1.5. When possibility theory is used in modeling, then the elements of the universe $\omega \in \Omega$ and subsets $A \subseteq \Omega$ represent states and groups of states of the world, possibility distribution values $\pi(\omega)$ represent the possibility of individual states, and possibility measure value $\Pi(A)$ represent the possibility of aggregates of states. Possibility values can be determined by the measurement procedures described in Chap. 4, but so far that allows only a static description. As discussed in Sec. 3.1.2, not only measurement, but also **possibilistic prediction** methods, here **possibilistic processes**, which modify and project possibilistic states forward in time, are necessary to complete the formulation of possibilistic models.

## 5.1 Mathematical Preliminaries

A number of mathematical ideas are necessary in order to proceed.

### 5.1.1 Semirings

Semirings are a common construct in abstract algebra, and have been used to ground general automata theory [92,205]. While we use them here on general sets, they can be extended to lattices [38].

**Definition 5.1 (Semiring)** Assume a **value set** $V = \{v\}$ with $0, 1 \in V$, and functions $\oplus\colon V^2 \mapsto V$ and $\otimes\colon V^2 \mapsto V$. Then $\mathcal{R} := \langle V, \oplus, \otimes, 0\rangle$ (denoted as $\langle V, \oplus, \otimes\rangle$, or just $\langle \oplus, \otimes\rangle$ where possible without confusion) is a **semiring** on $V$ if:

- $\langle V, \oplus, 0\rangle$ is an Abelian monoid ($\oplus$ is a commutative and associative operator with identity 0);
- $\langle V, \otimes\rangle$ is a semigroup ($\otimes$ is an associative operator).
- $\otimes$ distributes over $\oplus$.

**Definition 5.2 (Additive Semiring)** A semiring $\mathcal{R}$ is **additive** if $\oplus = +$.

**Corollary 5.3** $\langle \mathrm{I\!R}, +, \times\rangle$ is an additive semiring.

**Proof:** Follows from the closure of $+$ and $\times$ on $\mathrm{I\!R}$, and the distributivity of $\times$ over $+$. ∎

**Corollary 5.4** If $\mathcal{R}$ is additive and $\otimes = \sqcap \in \{\wedge, \times, \sqcap_m, \sqcap_w\}$, then $\otimes = \times$.

**Proof:** Counterexamples for the other cases are easily constructed. ∎

**Definition 5.5 (Conorm Semiring)** A semiring $\mathcal{R}$ is a **conorm semiring** if $\mathcal{R} = \langle [0,1], \sqcup, \sqcap\rangle$ for some conorm $\sqcup$ and norm $\sqcap$.

Note that $\langle +, \times\rangle$ is not (necessarily) a conorm semiring, because $+$ is not (generally) closed on $[0,1]$.

Of course $\langle \oplus, \otimes\rangle$ need not be a dual $\langle \sqcup, \sqcap\rangle$ pair to be a conorm semiring. The key property for $\langle \sqcup, \sqcap\rangle$ pairs is DeMorgan, while the key property for $\langle \oplus, \otimes\rangle$ semirings is distributivity. Still, some $\langle \sqcup, \sqcap\rangle$ pairs are also semirings.

**Theorem 5.6** $\langle [0,1], \vee, \sqcap\rangle$ is a conorm semiring for all norms $\sqcap$.

**Proof:** The first two conditions of the definition of semirings (5.1) are trivial. To prove distributivity, we need to show that

$$\forall x, y, z \in [0,1], \quad x \sqcap (y \vee z) = (x \sqcap y) \vee (x \sqcap z).$$

Let $y \leq z$. Then $x \sqcap (y \vee z) = x \sqcap z$. And from the corollary (2.6) in Sec. 2.1.2, $x \sqcap y \leq x \sqcap z$, so that $(x \sqcap y) \vee (x \sqcap z) = x \sqcap z$. An analogous result holds for $y \geq z$. ∎

**Proposition 5.7** If $\mathcal{R} = \langle [0,1], \sqcup, \sqcap\rangle$ is a conorm semiring and $\sqcup$ and $\sqcap$ are dual, then $\mathcal{R} = \langle [0,1], \vee, \wedge\rangle$.

**Proof:** Dubois and Prade [64, p. 80] have shown that $\vee$ and $\wedge$ are the sole distributive $\langle \sqcup, \sqcap\rangle$ dual pairs. ∎

### 5.1.2 Fuzzy Relation Composition

The composition of fuzzy relations is a standard subject in fuzzy theory, and treated in many standard textbooks [138,155] (see also Yeh and Bang [319] and Yager [313]). Here fuzzy matrix and relation composition is cast anew in terms of semirings.

In this subsection assume four finite universes of discourse $\Omega_1, \Omega_2, \Omega_3, \Omega_4$ where $n := |\Omega_1|, m := |\Omega_2|, p := |\Omega_3|, q := |\Omega_4|$.

**Definition 5.8 (Fuzzy Relation (Matrix) Composition)**

- Given two fuzzy relations $\tilde{R} \tilde{\subseteq} \Omega_1 \times \Omega_3, \tilde{S} \tilde{\subseteq} \Omega_3 \times \Omega_2$ in matrix form $R_{n\times p} = [R_{ik}], S_{p\times m} = [S_{kj}]$, and a conorm semiring $\mathcal{R} = \langle \sqcup, \sqcap \rangle$, then the **composition** of $R$ and $S$ through $\mathcal{R}$ is a fuzzy relation

$$\tilde{T} := \tilde{R} \circ_{\mathcal{R}} \tilde{S} \tilde{\subseteq} \Omega_1 \times \Omega_2$$

  where, in matrix form

$$T_{n\times m} := [T_{ij}], \qquad T_{ij} := \bigsqcup_{k=1}^{p} R_{ik} \sqcap S_{kj}.$$

- Generally $\mathcal{R}$ will be fixed, and let $\circ := \circ_{\mathcal{R}}$.

- For a square fuzzy matrix $R_{n\times n} \tilde{\subseteq} \Omega^2$, let $R^1 := R$, and for $2 \leq t \in \mathcal{W}$, let $R^t := R^{t-1} \circ R$.

- A square fuzzy matrix $R_{n\times n} \tilde{\subseteq} \Omega^2$ can also be denoted as an $n$-row vector of $n$-column vectors $R^{(j)} := \langle R_{1j}, R_{2j}, \ldots, R_{nj} \rangle^T$, for $1 \leq j \leq n$, so that

$$R = \left\langle R^{(j)} \right\rangle = \left\langle R^{(1)}, R^{(2)}, \ldots, R^{(n)} \right\rangle = \left[ \begin{pmatrix} R_{11} \\ R_{21} \\ \vdots \\ R_{n1} \end{pmatrix}, \begin{pmatrix} R_{12} \\ R_{22} \\ \vdots \\ R_{n2} \end{pmatrix}, \ldots, \begin{pmatrix} R_{1n} \\ R_{2n} \\ \vdots \\ R_{nn} \end{pmatrix} \right].$$

The proof of the following well-known result will be useful later.

**Corollary 5.9 (Matrix Composition Associativity)** [138] Fuzzy matrix composition is associative.

**Proof:** Assume three fuzzy matrices

$$R_{n\times p} = [R_{ik}] \tilde{\subseteq} \Omega_1 \times \Omega_3, \qquad S_{p\times q} = [S_{kl}] \tilde{\subseteq} \Omega_3 \times \Omega_4, \qquad T_{q\times m} = [T_{lj}] \tilde{\subseteq} \Omega_4 \times \Omega_2.$$

Then in virtue of the distributivity of $\sqcap$ over $\sqcup$, the commutivity of $\sqcup$ and $\sqcap$, and the associativity of $\sqcup$,

$$\begin{aligned}
(\widetilde{R} \circ (\widetilde{S} \circ \widetilde{T}))_{ij} &= \bigsqcup_k R_{ik} \sqcap \left( \bigsqcup_l S_{kl} \sqcap T_{lj} \right) \\
&= \bigsqcup_k \bigsqcup_l R_{ik} \sqcap S_{kl} \sqcap T_{lj} \\
&= \bigsqcup_l \bigsqcup_k R_{ik} \sqcap S_{kl} \sqcap T_{lj} \\
&= \bigsqcup_l \left( \bigsqcup_k R_{ik} \sqcap S_{kl} \right) \sqcap T_{lj} \\
&= ((\widetilde{R} \circ \widetilde{S}) \circ \widetilde{T})_{ij}.
\end{aligned}$$

■

## 5.2 General Processes

The mathematical formulation of general processes, based on semirings, will be derived first. All of the familiar classes of processes and automata follow as special cases, and possibilistic processes are defined as a new special case.

General processes have been treated in the context of category theory, for example by Eilenberg [80], Arbib and his colleagues [3, 4], Peeva and his colleagues [205, 288] and Ray and Chatterjee [234]. But the primary source for us here is again the work of Gaines and Kohout [91, 92] which was used extensively in Chap. 3. Our treatment is slightly more general and complete, however.

**Definition 5.10 (General Process)** A system $\mathcal{Z}^* := \langle \Omega, \mathcal{R}, \sigma, \phi^0 \rangle$ is a **general process** if:

- $\mathcal{R}$ is a semiring.
- $\sigma \colon \Omega^2 \mapsto V$ is the **transition function**.
- $\phi^t \colon \Omega \mapsto V$ is the **state function** where $\phi^0$ is given as an **initial state** and $\forall \omega_i \in \Omega, 1 \leq t \in \mathcal{W}$,

$$\phi^t(\omega_i) := \bigoplus_{\omega_j \in \Omega} \phi^{t-1}(\omega_j) \otimes \sigma(\omega_i, \omega_j). \tag{5.11}$$

**Definition 5.12 (Conorm Process)** A general process $\mathcal{Z}^*$ is a **conorm process** $\mathcal{Z}$ if $\mathcal{R}$ is a conorm semiring. Denote $\mathcal{Z} = \langle \sqcup, \sqcap, \sigma, \phi^0 \rangle$.

If not specified otherwise, then in the sequel processes will be presumed to be conorm processes, and the unqualified term "process" will be taken to mean conorm process. But it is understood that all conorm processes are general processes.

**Proposition 5.13** Because $\sqcup$ and $\sqcap$ are both closed on $[0,1]$, therefore for a conorm process $\mathcal{Z}$,

- $\phi^t$ is a fuzzy subset of $\Omega$, denoted $\tilde{\phi}^t \tilde{\subseteq} \Omega$, as in (2.34);
- $\sigma$ is a fuzzy relation on $\Omega$, denoted $\tilde{\sigma} \tilde{\subseteq} \Omega^2$.

So processes can be cast in terms of the composition of fuzzy relations.

**Definition 5.14 (Process, Matrix Form)** A process $\mathcal{Z}$ can be defined in matrix form where

- Letting $\sigma_{ij} := \sigma(\omega_i, \omega_j), 1 \le i,j \le n$, then $\sigma$ is taken as a matrix $\sigma_{n\times n} := [\sigma_{ij}]$.
- Letting $\phi_i^t := \phi^t(\omega_i)$, then $\phi^t$ is taken as a column vector

$$\vec{\phi}^t_{n\times 1} := \left\langle \phi_1^t, \phi_2^t, \ldots, \phi_n^t \right\rangle^T$$

  so that the transition function has the form of fuzzy matrix composition

$$\vec{\phi}^t = \vec{\phi}^{t-1} \circ \sigma = \begin{pmatrix} \phi_1^{t-1} \\ \phi_2^{t-1} \\ \vdots \\ \phi_n^{t-1} \end{pmatrix} \circ \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}.$$

The definition of a process is justified by the following argument:

1. $t$ indicates discrete time, where $t = 0$ is an initial condition.
2. $\phi^t(\omega_i) \in [0,1]$ represents the degree, extent, or certainty (generically here called "degree") that $\omega_i$ is the actual state at time $t$. The purpose of a process is to move $\phi$ forward in time.
3. $\sigma(\omega_i, \omega_j) \in [0,1]$ represents the (time independent) degree that a transition will be made from state $\omega_j$ to state $\omega_i$.
4. $\otimes$ combines a current degree with some degree of transition. It should be a norm because:
   - If either the current or transition degrees are maximal, then the combined degree can only be the other degree;

- The combined degree should increase with each of the combining degrees (monotonicity).

5. $\oplus$ aggregates combined degrees over all possible current states at one time. It should be a conorm because:

    - If any of the combined degrees are minimal, then they should not be factors in the aggregation of degrees;
    - The aggregated degrees should increase with each of the aggregating degrees (monotonicity).

6. The requirement that $\langle [0,1], \oplus, \otimes \rangle$ should be a semiring is a result of the necessity of distributivity in the proof of the associativity of fuzzy matrix composition (5.9), so that states can be projected forward arbitrarily in time by the corollary (5.15).

**Corollary 5.15** Given a process $\mathcal{Z}$, then $\vec{\phi}^t = \vec{\phi}^0 \circ \sigma^t$.

**Proof:** First, $\vec{\phi}^1 = \vec{\phi}^0 \circ \sigma = \vec{\phi}^0 \circ \sigma^1$. Also by fuzzy matrix composition associativity (5.9),

$$\vec{\phi}^t = \vec{\phi}^{t-1} \circ \sigma = (\vec{\phi}^{t-2} \circ \sigma) \circ \sigma = \vec{\phi}^{t-2} \circ (\sigma \circ \sigma) = \vec{\phi}^{t-2} \circ \sigma^2.$$

So by induction,

$$\vec{\phi}^t = \vec{\phi}^{t-t} \circ \sigma^t = \vec{\phi}^0 \circ \sigma^t.$$

■

Whereas the $\omega_i \in \Omega$ are the actual states of the system, as discussed in Sec. 3.1.4 the state function $\phi^t$ establishes a meta-state representation ("hyper-states" for Gaines and Kohout) for the process $\mathcal{Z}$. So where $\phi_i^t$ indicates uncertainty about $\omega_i$ at time $t$, the state function (5.11) establishes a deterministic, functional relation between $\phi^{t-1}$ and $\phi^t$.

## 5.3 Classes of Processes

A variety of specialized classes of processes are available. Most of these (deterministic, nondeterministic, and stochastic) are quite familiar in systems theory. Fuzzy processes and automata[1] have been used in fuzzy theory for a long time as well. Possibilistic processes will be introduced here as a new concept related to normalization

[1]This section is about processes only, and automata are not formally introduced until Sec. 5.6.1. Nevertheless, automata are built directly from processes, and are used more in both theory and applications than processes are. Therefore automata will be mentioned some in passing in this section.

criteria, and the role of normalization in distinguishing stochastic and possibilistic processes from the other classes will be considered.

### 5.3.1 Specialization Conditions

The various individual specializations that can applied to general and conorm processes are detailed in this section. Their combinations will be considered in the next section.

Certain concepts developed in Chap. 2 carry over to processes.

**Definition 5.16 (Crisp Process)** A state vector $\vec{\phi}^t$ is **crisp** if $\forall \omega_i \in \Omega, \phi_i \in \{0,1\}$. A process is crisp if $\forall t \geq 0, \vec{\phi}^t$ is crisp.

**Definition 5.17 (Certain Process)** A state vector $\vec{\phi}^t$ is **certain** if $\exists! \omega_i \in \Omega, \phi_i = 1$ and $\forall \omega_k \neq \omega_i, \phi_k = 0$. A process is certain if $\forall t \geq 0, \vec{\phi}^t$ is certain.

Of course, crisp and certain vectors and processes recall crisp and certain distributions from (2.25) and (2.97) respectively.

**Definition 5.18 (Subnormal and Normal Processes)** Assume a general process $\mathcal{Z}^*$.

- A state vector $\vec{\phi}^t$ is **subnormal** if $\bigoplus_i \phi_i^t \leq 1$, and **normal** if equality holds.
- $\mathcal{Z}^*$ is subnormal (resp. normal) if $\forall t \geq 0, \vec{\phi}^t$ is subnormal (normal).
- $\mathcal{Z}^*$ is **transition subnormal** (resp. **transition normal**) if $\forall 1 \leq j \leq n$ the column vector $\sigma^{(j)}$ is subnormal (normal).

Normalization links processes from fuzzy theory into information theory through distributions. The following was first shown by Gaines and Kohout, but is re-proven here because of its significance and the slightly different concepts used.

**Theorem 5.19 (Process Normalization)** [92] Given a process $\mathcal{Z}$ where $\vec{\phi}^0$ is normal and $\mathcal{Z}$ is transition normal, then $\mathcal{Z}$ is normal.

**Proof:** Proof by induction. First, $\vec{\phi}^0$ is normal by premise. Then, let $t > 0$ be fixed, assume $\vec{\phi}^{t-1}$ is normal, and denote $\phi_j := \phi_j^{t-1}$ and $xy := x \sqcap y$. Then in virtue of the commutivity of $\sqcup$ and $\sqcap$ and the distributivity of $\sqcap$ over $\sqcup$

$$\bigsqcup_i \phi_i^t = \bigsqcup_i \bigsqcup_j \phi_j \sigma_{ij} = \bigsqcup_j \bigsqcup_i \phi_j \sigma_{ij} = \bigsqcup_j \phi_j \bigsqcup_i \sigma_{ij}.$$

Now because $\mathcal{Z}$ is transition normal, $\forall j, \bigsqcup_i \sigma_{ij} = 1$, therefore

$$\bigsqcup_i \phi_i^t = \bigsqcup_j \phi_j 1,$$

so that finally from the identity of 1 for $\sqcap$ and the normality of $\phi$,

$$\bigsqcup_i \phi_i^t = \bigsqcup_j \phi_j = 1.$$

■

Certainty imposes the highest constraint, and is thus a very special case.

**Corollary 5.20** Given a process $\mathcal{Z}$, then if a state vector $\vec{\phi}^t$ (resp. $\mathcal{Z}$ itself) is certain, then $\vec{\phi}^t$ (resp. $\mathcal{Z}$) is both crisp and normal.

**Proof:** Proof will be made for a single state vector only. The result for $\mathcal{Z}$ follows from the result for all state vectors. Let $\vec{\phi}^t$ be certain. Its crispness is obvious from the definitions (5.16) and (5.18). Its normality follows from the identity of 0 for $\sqcup$

$$\bigsqcup_i \phi_i^t = 0 \sqcup 0 \sqcup \cdots \sqcup 1 \sqcup \cdots \sqcup 0 = 1.$$

■

The choice of conorm aggregation operator is also a consideration in characterizing a conorm process.

**Definition 5.21 (General Fuzzy Process)** A process $\mathcal{Z}$ is a **general fuzzy process** if $\sqcup = \vee$.

Note that general fuzzy processes are conorm processes, not general processes.

From the corollary (5.6), $\vee$ forms a semiring with any norm $\sqcap$. So general fuzzy processes are defined on semirings of the form $\langle \vee, \sqcap \rangle$, and the state function takes the form

$$\phi_i^t = \bigvee_{j=1}^n \phi_j^{t-1} \sqcap \sigma_{ij}.$$

**Definition 5.22 (Proper Fuzzy Process)** A general fuzzy process $\mathcal{Z}$ is a **proper fuzzy process** if $\mathcal{R} = \langle \vee, \wedge \rangle$.

In proper fuzzy processes the state function is

$$\phi_i^t = \bigvee_{j=1}^n \phi_j^{t-1} \wedge \sigma_{ij}.$$

$\wedge$ is far and away the favored norm used in fuzzy processes. Whereas $\langle \vee, \times \rangle$ has been used somewhat, for example by Santos [249,250], little attention has been paid to the use of other norms (see Sec. 5.5.2.1).

**Definition 5.23 (Additive General Process)** A general process $\mathcal{Z}^*$ is **additive** if $\mathcal{R}$ is an additive semiring, that is, if $\oplus = +$.

Additive processes move toward the standard form of matrix composition

$$\vec{\phi}^t = \vec{\phi}^{t-1} \cdot \sigma, \qquad \phi_i^t = \sum_{j=1}^{n} \phi_j^{t-1} \otimes \sigma_{ij}.$$

Finally there is a degenerate case.

**Definition 5.24 (Degenerate Process)** A state vector $\vec{\phi}^t$ is **degenerate** if $\forall \omega_i \in \Omega, \phi_i = 0$. A process is degenerate if $\forall t \geq 0, \vec{\phi}^t$ is degenerate.

**Corollary 5.25 (Degeneracy Conditions)** Given a process $\mathcal{Z}$, if $\exists t_0, \vec{\phi}^{t_0}$ is degenerate, then $\forall t > t_0, \vec{\phi}^t$ is degenerate.

**Proof:** Fix $t_0$, and let $\vec{\phi}^{t_0}$ be degenerate, so that $\forall i, \phi_i^{t_0} = 0$. Then $\forall i, j, \phi_j^{t_0} \sqcap \sigma_{ij} = 0$, so that $\forall i, \phi_i^{t_0+1} = 0 \sqcup \cdots \sqcup 0 = 0$. The result follows by induction. ■

**Corollary 5.26** Given a process $\mathcal{Z}$, if $\vec{\phi}^0$ is degenerate then $\mathcal{Z}$ is degenerate.

**Proof:** Trivial from the definition (5.24) and the corollary (5.25). ■

### 5.3.2 Combinations of Specializations

In this section all sixteen possible cases of processes, resulting from all the combinations of the conditions introduced in Sec. 5.3.1 (crisp and not crisp, certain and not certain, additive and general fuzzy, and normal and not necessarily normal), will be characterized. This consideration leads not only to the identification of the common classes from automata theory, but also an important new class related strictly to possibility theory.

#### 5.3.2.1 Deterministic Processes

Deterministic processes and automata are very familiar in systems theory and discrete mathematics generally (Hopcroft and Ullman [122] and Starke [273] are standard references). They simply follow a mapping of $\Omega$ to itself from some initial state.

**Definition 5.27 (Deterministic Process)** [122] A **deterministic process** is a system $\mathcal{Z}_d := \langle \Omega, \delta, \omega^* \rangle$ where $\delta \colon \Omega \mapsto \Omega$ is the **next state function** and $\omega^* \in \Omega$ is the **initial state**.

**Theorem 5.28 (Deterministic Process Conditions)** The classes of deterministic and certain processes are equivalent.

**Proof:**

1. Assume a certain process $\mathcal{Z}$, and denote $\omega^t$ as the unique $\omega_i$ for which $\phi_i^t = 1$. Then construct $\mathcal{Z}_d$ by letting

$$\omega^* := \omega^0, \qquad \delta(\omega^*) = \delta(\omega^0) := \omega^1, \qquad \delta(\delta(\omega^*)) = \delta(\omega^1) := \omega^2,$$

   and generally

$$\delta^t(\omega^*) = \delta(\omega^{t-1}) := \omega^t$$

   until a cycle is achieved. Finally, for any $\omega_i$ for which $\nexists t > 0, \omega^t = \omega_i$, let $\delta(\omega_i) \in \Omega$ arbitrarily. The construction is valid on its face.

2. Assume a deterministic process $\mathcal{Z}_d$, and denote $\vec{\phi}(\omega_i)$ as the column vector for which $i = j \rightarrow \phi(\omega_i)_j = 1$ and $i \neq j \rightarrow \phi(\omega_i)_j = 0$. To construct $\mathcal{Z}$, first let $\vec{\phi}^0 = \vec{\phi}(\omega^*)$. Then for $1 \leq j \leq n$, let $\sigma^{(j)} = \vec{\phi}(\delta(\omega_j))$. Finally, $\mathcal{R}$ can be any conorm semiring. To verify the construction, denote $\omega^t = \delta^t(\omega^*)$, and fix $t$. Now $\phi_j^{t-1} = 1$ only if $\omega^{t-1} = \omega_j$, and $\forall i, \sigma_{ij} = 1$ only if $\delta(\omega_j) = \omega_i$. Therefore $\phi_i^t = \bigsqcup_j \phi_j^{t-1} \sqcap \sigma_{ij} = 1$ only if $\omega^t = \delta(\omega^{t-1})$.

■

#### 5.3.2.2 Stochastic Processes

Stochastic processes and automata are also a staple of systems theory, and for years have been the standard representation of machines which operate under conditions of uncertainty (see, for example, texts by Carlyle [26], Doberkat [49], Paz [203], and Starke [273], and descriptions of applications by Glorioso and Osorio [99] and Grossing and Zeilinger [106].)

**Definition 5.29 (Stochastic Process)** [203] A **stochastic process** is a system $\mathcal{Z}_p := \left\langle \Omega, \mathbf{P}, \vec{p}^0 \right\rangle$, where

- The matrix $\mathbf{P} = \left\langle \mathbf{P}^{(j)} \right\rangle = [\mathbf{P}_{ij}]$, so that $\forall j, \mathbf{P}^{(j)}$ is the vector representation of a conditional probability distribution function $p(\cdot|\omega_j) \colon \Omega \mapsto [0,1]$ and $\mathbf{P}_{ij} := p(\omega_i|\omega_j)$ with $\sum_i p(\omega_i|\omega_j) = 1$;

- $\vec{p}^t$ is a probability distribution on $\Omega$ with, $\forall t > 0$,

$$\vec{p}^t := \vec{p}^{t-1} \cdot \mathbf{P}, \qquad p_i^t := \sum_j p_j^{t-1} \times p(\omega_i|\omega_j).$$

$p_i^t = p^t(\omega_i)$ is the probability of being in state $\omega_i$ at time $t$, and $\mathbf{P}_{ij} = p(\omega_i|\omega_j)$ is the probability of transiting from state $\omega_j$ to $\omega_i$.

**Theorem 5.30 (Stochastic Process Conditions)** A stochastic process $\mathcal{Z}_p$ is an additive conorm process.

**Proof:** Assume a stochastic process $\mathcal{Z}_p$. Then let

$$\mathcal{Z} := \left\langle \Omega, \langle [0,1], \sqcup_m, \times \rangle, \mathbf{P}, \vec{p}^{\,0} \right\rangle.$$

From the definitions of additive general processes (5.23) and conorm processes (5.2), all we need to show is that $\langle [0,1], \sqcup_m, \times \rangle$ acts as an additive conorm semiring on the matrix $\mathbf{P}$ and the state vector $\vec{p}^{\,0}$.

1. From the definition of $\sqcup_m$ in Table 2.1, it is obvious that

$$\forall x, y \in [0,1], \quad x + y \leq 1 \rightarrow x \sqcup_m y = (x+y) \wedge 1 = x + y.$$

   Therefore

$$\forall t \geq 0, \quad \sum_i \phi_i^t \leq 1 \rightarrow \sum_i \phi_i^t = \bigsqcup_{i=1}^{n}{}_m \, \phi_i^t.$$

2. Because $\vec{p}^{\,0}$ is a probability distribution, it is normal under $+$, so that from the above argument $\sum_i \phi_i^0 = \bigsqcup_{i=1}^{n}{}_m \, \phi_i^0$, and therefore $\vec{p}^{\,0}$ is normal under $\sqcup_m$.

3. Because $\forall j, \mathbf{P}^{(j)}$ is a probability distribution conditional on $\omega_j$, therefore $\mathbf{P}$ is transition normal under $+$, and so it is also under $\sqcup_m$.

4. Because $\times$ is a norm which distributes over $+$, therefore when $\sqcup_m = +$ then $\langle \sqcup_m, \times \rangle = \langle +, \times \rangle$ is a conorm semiring.

5. So at time $t = 0$, $\mathcal{Z}$ is a conorm process satisfying the preconditions for process normalization (5.19). Therefore by the argument used in that proof, $\phi^1$ is also normal under $\sqcup_m$. So by induction, $\forall t \geq 0, \phi^t$ is normal under $\sqcup_m$, so $\mathcal{R}$ is a conorm semiring for general $t$.

■

Restricting $\oplus = \sqcup_m = +$ forces $\oplus = +$ to be closed on $[0,1]$, and thus a conorm. Then transition normalization forces the transition matrix $\sigma$ to be a collection of conditional probability distributions, one for each of the column vectors $\sigma^{(j)}$. So finally, when $\phi^0$ is also additively normal, then $\langle +, \times, \sigma, \phi^0 \rangle$ becomes an additive conorm process, and the state function assumes the familiar form

$$p_i^t = \sum_{j=1}^{n} p_j^{t-1} \times p(\omega_i|\omega_j).$$

$\mathcal{Z}$ is now a Markov process of degree 1 [203, pp. 67 ff.].

So while $\langle +, \times \rangle$ is not usually a conorm semiring, it is when $\langle +, \times \rangle$ composition is used *in probability theory*, so that $+$ is restricted to $\sqcup_m$. This result links traditional stochastic processes into the world of systems and processes based on GIT, and shows that, like probability theory itself, stochastic systems are an important, but single, component of an overall systems theory including GIT.

There is one last case of additive processes to consider.

**Corollary 5.31** If a normal additive process $\mathcal{Z}$ is crisp, then $\mathcal{Z}$ is deterministic.

**Proof:** Fix $t$. Since $\mathcal{Z}$ is crisp, if $\exists i_1, i_2$ with $p^t_{i_1} > 0 < p^t_{i_2}$, then $p^t_{i_1} = p^t_{i_2} = 1$, and so $\sum_i p^t_i > 1$, which it cannot since $\mathcal{Z}$ is normal. If $\forall i, p^t_i = 0$, then $\sum_i p^t_i = 0$, which it also cannot since $\mathcal{Z}$ is normal. Therefore $\mathcal{Z}$ must be certain, so that from (5.28) $\mathcal{Z}$ is deterministic. ∎

#### 5.3.2.3 Nondeterministic Processes

Another common class of automata is based on nondeterministic processes, which are used especially in computer science, algorithm theory, and mathematical linguistics. Again, Hopcroft and Ullman [122] and Starke [273] are standard references.

**Definition 5.32 (Nondeterministic Process)** [122] A **nondeterministic process** is a system $\mathcal{Z}_n := \langle \Omega, \delta, A^0 \rangle$ where $\delta \colon \Omega \mapsto 2^\Omega$ is the **transition function**, $A^0 \subseteq \Omega$ is the **initial state**, and the **state** is

$$\forall t > 0, \quad A^t := \bigcup_{\omega_i \in A^{t-1}} \delta(\omega_i).$$

Whereas a deterministic process follows an initial state through a state mapping, a nondeterministic process follows an initial *subset* of states through a state *subset* mapping.

**Theorem 5.33 (Nondeterministic Process Conditions)** The classes of crisp, non-certain, normal, general fuzzy processes and nondeterministic processes are equivalent.

**Proof:**

1. Assume a crisp, normal, general fuzzy process $\mathcal{Z}$. Let $A(\vec{\phi})$ be the set determined by interpreting $\phi$ as a characteristic function, so that if $\phi_i = 1$ then $\omega_i \in A(\vec{\phi})$ and if $\phi_i = 0$ then $\omega_i \notin A(\vec{\phi})$. Then construct $\mathcal{Z}_n$ by first letting

$A^0 := A(\vec{\phi}^0)$, and then, $\forall \omega_j \in \Omega$, letting $\delta(\omega_j) := A(\sigma^{(j)})$. To verify this construction, fix $t$ and $i$. Then

$$\phi_i^t = 1 \to \bigsqcup_j \phi_j^{t-1} \sqcap \sigma_{ij} = 1.$$

This will only be the case if $\exists j, \phi_j^{t-1} = 1$ and $\sigma_{ij} = 1$. Similarly, $\omega_i \in A^t$ if $\omega_i \in \bigcup_{\omega_j \in A^{t-1}} \delta(\omega_j)$, or in other words if $\exists j, \omega_j \in A^{t-1}$ and $\omega_i \in \delta(\omega_j)$.

2. Assume a nondeterministic process $\mathcal{Z}_n$. Now let $\mathcal{Z} := \langle \vee, \sqcap, \sigma, \chi_{A^0} \rangle$ for some norm $\sqcap$, recalling that $\chi_A$ is the characteristic function of $A \subseteq \Omega$, and $\forall 1 \leq j \leq n$, letting $\sigma^{(j)} := \chi_{\delta(\omega_j)}$. To verify the construction, fix $t$. Then $\omega_i \in A^t$ if $\omega_i \in \bigcup_{\omega_j \in A^{t-1}} \delta(\omega_j)$, so that $\exists \omega_{j_0} \in A^{t-1}, \omega_i \in \delta(\omega_{j_0})$. This will only be the case if $\phi_{j_0}^{t-1} = 1$ (so that $\omega_{j_0} \in A^{t-1}$), and $\sigma_{ij_0} = 1$ (so that $\omega_i \in \delta(\omega_{j_0})$). Then $\phi_{j_0}^{t-1} \sqcap \sigma_{ij_0} = 1$, so that $\phi_i^t = \bigvee_j \phi_j^{t-1} \sqcap \sigma_{ij} = 1$, so that $\omega_i \in A^t$.

■

**Corollary 5.34** A crisp, subnormal, general fuzzy process $\mathcal{Z}$ is degenerate.

**Proof:** Let $\vec{\phi} := \vec{\phi}^0$. Since $\mathcal{Z}$ is crisp, $\forall i, \phi_i \in \{0,1\}$. But subnormality of a general fuzzy process requires that $\bigvee_i \phi_i < 1$, so that $\forall i, \phi_i = 0$. The result then follows from the second degeneracy corollary (5.26). ■

#### 5.3.2.4 Possibilistic Processes

So far in this section we have considered all crisp, additive, and certain processes. The only two classes left are non-crisp, non-certain, general fuzzy processes, both normal and subnormal.

So finally, we arrive at the definition and characterization of **possibilistic processes**.

**Definition 5.35 (Possibilistic Process)** A **possibilistic process** is a system $\mathcal{Z}_\pi := \langle \Omega, \sqcap, \boldsymbol{\Pi}, \vec{\pi}^0 \rangle$ for some norm $\sqcap$, where:

- The matrix $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_{ij}] = \left\langle \boldsymbol{\Pi}^{(j)} \right\rangle$, so that $\forall j, \boldsymbol{\Pi}^{(j)}$ is the vector representation of a conditional possibility distribution function $\pi(\cdot|\omega_j) \colon \Omega \mapsto [0,1]$ and $\boldsymbol{\Pi}_{ij} := \pi(\omega_i|\omega_j)$ with
$$\bigvee_i \pi(\omega_i|\omega_j) = 1; \qquad (5.36)$$
- $\vec{\pi}^t$ is a possibility distribution on $\Omega$ with, $\forall t > 0$,
$$\vec{\pi}^t := \vec{\pi}^{t-1} \circ \boldsymbol{\Pi}, \qquad \pi_i^t := \bigvee_j \pi_j^{t-1} \sqcap \pi(\omega_i|\omega_j).$$

$\pi_i^t = \pi^t(\omega_i)$ is the possibility of being in state $\omega_i$ at time $t$, and $\boldsymbol{\Pi}_{ij} = \pi(\omega_i|\omega_j)$ is the possibility of transiting from state $\omega_j$ to $\omega_i$. Conditional possibility will be fully discussed in Sec. 5.5.3, and (5.36) will be proved with corollary (5.51).

**Theorem 5.37 (Possibilistic Process Conditions)** A possibilistic process $\mathcal{Z}_\pi$ is a normal general fuzzy process.

**Proof:** Assume a possibilistic process $\mathcal{Z}_\pi$. Then let $\mathcal{Z} := \langle \vee, \sqcap, \boldsymbol{\Pi}, \vec{\pi}^0 \rangle$.

1. Clearly by the definition (5.21) $\mathcal{Z}$ is a general fuzzy process.

2. $\vec{\pi}^0$ is normal under $\vee$ because it is a possibility distribution. Then from (5.36) and the definition (5.18), $\mathcal{Z}$ is transition normal under $\vee$. So from process normalization (5.19), $\mathcal{Z}$ is normal.

■

#### 5.3.2.5 Fuzzy Processes Proper

This is in fact the end of the specification of classes of processes. The remaining case of non-crisp, non-certain, subnormal, general fuzzy processes are simply general fuzzy processes from their definition (5.21). These encompass the general case of processes which use $\langle \vee, \sqcap \rangle$ composition, but which are not necessarily normal.

General fuzzy automata, based on general fuzzy processes, were introduced and examined in the late 1960's in a series of papers by Santos [246], Santos and Wee [252], and Wee and Fu [302]. Properties of fuzzy automata (and languages based on them [175, 186, 247]) were investigated (see for example Dubois and Prade [55], Kandel and Lee [135], Santos [248], and Pedrycz [204]), and some applications where made, for example in pattern recognition [285] and parameter optimization [308]. Fuzzy automata and languages have received considerably less attention since the 1980's. Some theoretical progress continues, for example by Močkor [189], but fuzzy automata get only a very short consideration in recent systems theory texts (for example, Dougherty and Giardina [50]).

It should also be noted that, because not only does $\wedge$ distribute over $\vee$, but also $\vee$ distributes over $\wedge$, that so-called **optimistic fuzzy automata** are defined on the semiring $\langle [0,1], \wedge, \vee \rangle$ when the ordering $\leq$ on $[0,1]$ is reversed to $\geq$ [252].

### 5.3.3 Relations Among the Classes

The relations among these concepts are not obvious. Table 5.1 summarizes the sixteen possible combinations detailed in Sec. 5.3.2. Each entry includes a description of the class and a reference to the definition or theorem which establishes it. A $\times$

| | | General Fuzzy: $\oplus = \vee$ | | Additive: $\oplus = \sqcup_m$ | |
|---|---|---|---|---|---|
| | | Normal | General | Normal | General |
| Crisp | Certain | Deterministic (5.28) | $\times$ (5.20) | Deterministic (5.28) | $\times$ (5.20) |
| | Not certain | Nondeterministic (5.33) | Degenerate (5.34) | Deterministic (5.31) | Additive General (5.23) |
| Not crisp | Certain | $\times$ (5.20) | $\times$ (5.20) | $\times$ (5.20) | $\times$ (5.20) |
| | Not certain | Possibilistic (5.37) | General fuzzy (5.21) | Stochastic (5.30) | Additive General (5.23) |

Table 5.1: Classes of processes and automata.

mark indicates that this possibility does not exist, with a reference to the theorem which determines this.

Not all of the different cases are mutually exclusive. For example, the possibilistic process conditions (5.37) establishes that a possibilistic process is a normal general fuzzy process. Nothing is said about the status of its crispness or certainty. Therefore crisp normal general fuzzy processes, that is nondeterministic processes from (5.33), are also possibilistic processes.

The various implication relations and subclasses are summarized in Fig. 5.1. The nodes are labels for the most relevant descriptions of each of the primary classes of processes. The arcs are labeled by the specializations made from general processes to reach the different classes. Some of the implications of the diagram will be discussed below in Sec. 5.5.2.

## 5.4 Properties of Possibilistic Processes

Some properties of possibilistic processes are now discussed. First, it should be noted that both the stochastic (5.30) and possibilistic process conditions (5.37) establish only sufficient, and not necessary, conditions. That is, in general, state vector normalization does not require transition normalization, and the converse of (5.19) does not generally hold.

In this section, denote the following:

- $t$ will be fixed, and denote $\vec{\pi} := \vec{\pi}^t$;
- Denote the next-state vector as $\vec{\pi}' := \vec{\pi}^{t+1} = \boldsymbol{\Pi} \circ \pi$.

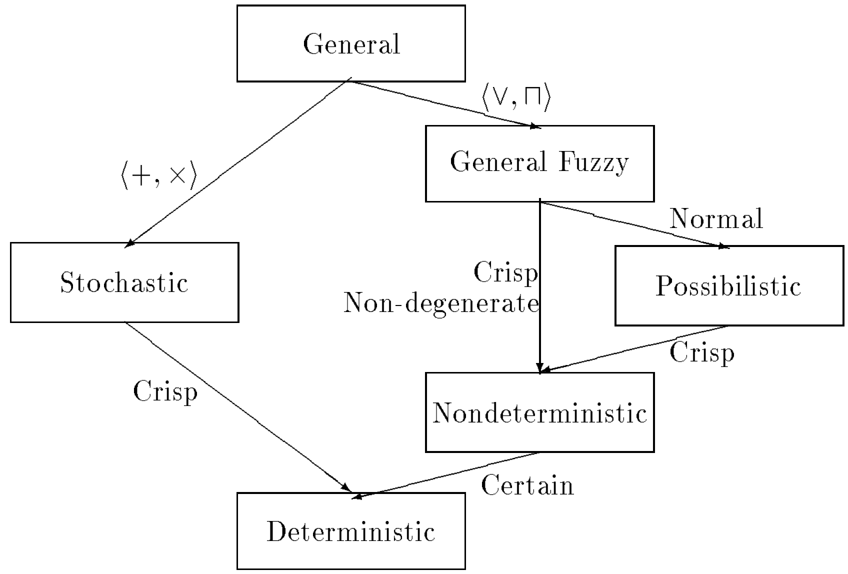


Figure 5.1: Relations among classes of processes

- For each row $i$ of $\boldsymbol{\Pi}$, let $\Theta(i) := \{\theta^i_k\}$ be the set of indices (if any) for which $\boldsymbol{\Pi}_{i,\theta^i_k} = 1$, where $1 \leq \theta^i_k \leq n$ and $1 \leq k \leq |\Theta(i)| \leq n$.
- Let $\theta(i)$ be the unique such index in each row $i$, if it exists.

The actions of possibilistic processes are dependent on a maximally normalized initial state vector $\pi^0$ and a transition normal conditional matrix $\boldsymbol{\Pi}$, and thus on the presence of 1's in the columns of $\boldsymbol{\Pi}$. The following lemma will be useful.

**Lemma 5.38** Let

$$R_{n\times p} := [R_{ik}] \,\tilde{\subseteq}\, \Omega_1 \times \Omega_3, \qquad S_{p\times m} := [S_{kj}] \,\tilde{\subseteq}\, \Omega_3 \times \Omega_2, \qquad T_{n\times m} := [T_{ij}] \,\tilde{\subseteq}\, \Omega_1 \times \Omega_2,$$

be fuzzy matrices, let $\mathcal{R} = \langle \vee, \sqcap \rangle$, and let $R \circ S = T$. If $\exists R_{i^*k^*} = 1$, then $\forall j, T_{i^*j} \geq S_{k^*j}$.

**Proof:** Assume $R_{i^*k^*} = 1$. Then $\forall j$,

$$\begin{aligned} T_{i^*j} &= \bigvee_{k=1}^{p} R_{i^*k} \sqcap S_{kj} \\ &= \left( \bigvee_{\substack{1\leq k\leq n \\ k\neq k^*}} R_{i^*k} \sqcap S_{kj} \right) \vee (R_{i^*k^*} \sqcap S_{k^*j}) = \left( \bigvee_{\substack{1\leq k\leq n \\ k\neq k^*}} R_{i^*k} \sqcap S_{kj} \right) \vee S_{k^*j} \\ &\geq S_{k^*j}. \end{aligned}$$

∎

Whereas each column of $\boldsymbol{\Pi}$ is guaranteed to have at least one unitary value, this is not necessarily the case for the rows. The column positions of the unitary elements in a row determine which elements of the state vector will provide lower bounds for the corresponding element of the next-state vector. If there are multiple unitary elements in a row, then the bound will be the maximum of all corresponding state values.

**Theorem 5.39** $\forall i$, if $\Theta(i) \neq \emptyset$, then

$$\pi'_i \geq \bigvee_{\theta^i_k \in \Theta(i)} \pi_{\theta^i_k}.$$

**Proof:** $\forall i, \forall \theta^i_k \in \Theta(i), \boldsymbol{\Pi}_{i,\theta^i_k} = 1$. Under the assignments $R_{n\times n} = \boldsymbol{\Pi}, S_{n\times 1} = \pi, T_{n\times 1} = \pi'$, then by the lemma (5.38), $\forall i, \forall \theta^i_k \in \Theta(i), \pi'_i \geq \pi_{\theta^i_k}$, and so the conclusion follows. ∎

**Corollary 5.40** $\forall i$, if $\theta(i)$ exists then $\pi'_i \geq \pi_{\theta(i)}$.

**Proof:** Follows directly from the theorem (5.39) under the assumption that $\Theta(i) = \{\theta(i)\}$. ∎

If each row of $\boldsymbol{\Pi}$ has a distinct and unique unitary element, then $\theta(i)$ exists and is an injective function creating a cyclic group permuting the elements of $\Omega$ through at most $n$ steps. Similarly, the bounds of the state vector will cyclically permute through a maximum of the $n$ elements of $\pi$.

**Theorem 5.41** If $\forall i, \theta(i)$ exists, and $\forall i_1 \neq i_2, \theta(i_1) \neq \theta(i_2)$, then $\forall i$ there is an index $\beta(i), 1 \leq \beta(i) \leq n$ such that $\pi^{\beta(i)}_i \geq \pi^0_i$.

**Proof:** Let $\lceil x \rceil$ be any value such that $\lceil x \rceil \in [x, 1]$, and let $\pi^0 = \langle \pi_1, \pi_2, \ldots, \pi_n \rangle^T$. By the corollary (5.40), the action of the automaton in the first step affects the transformation

$$\pi^0 \mapsto \pi^1 = \left\langle \lceil \pi_{\theta(1)} \rceil, \lceil \pi_{\theta(2)} \rceil, \ldots, \lceil \pi_{\theta(n)} \rceil \right\rangle^T.$$

The next step of the automaton affects the further transformation

$$\pi^1 \mapsto \pi^2 = \left\langle \lceil \pi_{\theta^2(1)} \rceil, \lceil \pi_{\theta^2(2)} \rceil, \ldots, \lceil \pi_{\theta^2(n)} \rceil \right\rangle^T.$$

In general,

$$\pi^t = \left\langle \lceil \pi_{\theta^t(1)} \rceil, \lceil \pi_{\theta^t(2)} \rceil, \ldots, \lceil \pi_{\theta^t(n)} \rceil \right\rangle^T.$$

Since $\forall i, \exists \beta(i), 1 \leq \beta(i) \leq n, \theta^{\beta(i)}(i) = i$, therefore

$$\forall i, \quad \pi_i^{\beta(i)} = \lceil \pi_{\theta^{\beta(i)}(i)} \rceil = \lceil \pi_i \rceil \geq \pi_i.$$

■

The placement of a unitary value on the diagonal of $\boldsymbol{\Pi}$ guarantees a monotonic increase of the corresponding state vector value.

**Corollary 5.42** If $\exists i, \theta(i) = i$, then $\pi'_i \geq \pi_i$.

**Proof:** Follows immediately from the corollary (5.40). ■

Finally, if $\boldsymbol{\Pi}$ has unitary values on all diagonal elements, then a result of Pedrycz [204] is recovered.

**Corollary 5.43** If $\forall i, \theta(i) = i$, then $\pi' \geq \pi$.

**Proof:** Follows immediately from the application of the corollary (5.42) to all columns. ■

## 5.5 Discussion

Possibilistic processes are defined analogously to stochastic processes, but replacing probabilistic with possibilistic mathematics, and thus the additive operator $+$ with $\vee$. As we have shown in Chap. 2, possibilistic normalization is the key criteria distinguishing possibility distributions from general fuzzy sets (just as additive normalization distinguishes probability distributions from general fuzzy sets). Similarly, whereas processes which use the $\vee$ operator are already available as fuzzy processes, possibilistic processes are distinct in virtue of their possibilistic normalization.

This section draws a number of conclusions about the nature of possibilistic processes, their relation to the other classes of processes, and to the rest of GIT and modeling within GIT.

### 5.5.1 Examples

First, some simple examples of possibilistic processes on a universe $\Omega = \{x, y, z\}$ are presented.

1. Consider the state transition diagram shown in Fig. 5.2. Each node is a state $\omega_i$. The arcs indicate state transitions, each of which is non-additively weighted

with the conditional possibility of the transition. So, for example, $\pi(x|y) = .8$, and the entire transition matrix is

$$\boldsymbol{\Pi} = \begin{bmatrix} 0.0 & 0.8 & 0.0 \\ 1.0 & 0.0 & 0.0 \\ 0.2 & 1.0 & 1.0 \end{bmatrix}.$$

$\boldsymbol{\Pi}$ is transition normal, with a 1 in each column.

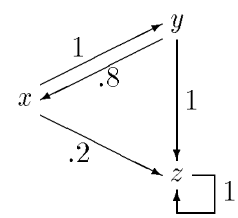


Figure 5.2: Weighted state transition diagram for a possibilistic process.

Possibilistic transition normalization requires that each node have an out-arc labeled 1. Nodes may have multiple out-arcs, however, so that a unitary weight does not indicate a *necessary*, but rather a *completely possible* transition. For example, from state $x$, transition to $z$ is .2 possible, but transition to $y$ is also (completely) possible. $z$ is an absorbing state, with self-transition the only possibility.

Now let $\sqcap = \wedge$, and assume that at time $t = 0$ the system begins *definitely* in state $x$. Then $\vec{\pi}^0 = \langle 1, 0, 0 \rangle^T$, a certain state vector, which is, of course, normal. Then the state vector at $t = 1$ is determined by

$$\vec{\pi}^1 = \vec{\pi}^0 \circ \boldsymbol{\Pi} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \circ \begin{bmatrix} 0.0 & 0.8 & 0.0 \\ 1.0 & 0.0 & 0.0 \\ 0.2 & 1.0 & 1.0 \end{bmatrix}.$$

so that

$$\begin{aligned} \pi_1^1 &= (1 \wedge 0) \vee (0 \wedge .8) \vee (0 \wedge 0) = 0, \\ \pi_2^1 &= (1 \wedge 1) \vee (0 \wedge 0) \vee (0 \wedge 0) = 1, \\ \pi_3^1 &= (1 \wedge .2) \vee (0 \wedge 1) \vee (0 \wedge 1) = .2. \end{aligned}$$

$\vec{\pi}^1$ is not a certain vector, so the system is now *not* unequivocally in one state. While it is completely possible that $y$ is the system state, $z$ is also .2 possible of being the system state, while $x$ is impossible.

Other future state vectors (and their nonspecificity values) are

| $t$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\vec{\pi}^t$ | 1.0 | 0.0 | 0.8 | 0.0 | 0.8 |
| | 0.0 | 1.0 | 0.0 | 0.8 | 0.0 |
| | 0.0 | 0.2 | 1.0 | 1.0 | 1.0 |
| $\mathbf{N}(\vec{\pi}^t)$ | 0.0 | 0.2 | 0.8 | 0.8 | 0.8 |

Since $\pi^2 = \pi^4$, the cycle will repeat. This is completely in keeping with the periodic behavior of proper fuzzy $\langle \vee, \wedge \rangle$ matrix composition, as shown by Thomason [285]. He has also shown [286] some sufficient conditions for the convergence of proper fuzzy processes.

Process normalization means that at each time, there is one state that is completely possible, and at various times that state can be any of $x, y$ or $z$. The process settles into a normal element of state $z$, which acts as a kind of attractor or absorbing state. But nevertheless, in virtue of the cycle between $x$ and $y$ with transition possibilities of 1 and .8 respectively, each of those states in turn remains .8 possible. But state $y$ can never recover the 1 possibility it has at $t = 1$, because the system is never again unequivocally in state $x$.

Note that no new elements are introduced in the sequence of the $\pi_i^t$. In other words,

$$\forall t > 0, \quad \forall i, \quad \pi_i^t \in \{\pi_i^0\} \cup \{\boldsymbol{\Pi}_{ij}\}.$$

This is because both $\vee$ and $\wedge$ are "conservative", in the sense that

$$\forall x, y \in \mathbb{R}, \qquad x \vee y \in \{x, y\}, \qquad x \wedge y \in \{x, y\}.$$

Note also how the theorems and corollaries of Sec. 5.4 is adhered to. In particular,

$$\Theta(1) = \emptyset, \qquad \Theta(2) = \{1\}, \quad \theta(2) = 1, \qquad \Theta(3) = \{2, 3\},$$

and so, for example, by theorem (5.39),

$$\pi_3^2 = 1 \geq \bigvee_{\theta \in \Theta(3)} \pi_\theta^1 = \pi_2^1 \vee \pi_3^1 = 1 \vee .2 = 1,$$

and by corollary (5.40),

$$\pi_2^1 = 1 \geq \pi_{\theta(2)}^0 = \pi_1^0 = 1.$$

Finally, the nonspecificity values of the state vectors are shown. In this case they increase monotonically, but that is by no means necessary.

2. For a slightly more complex example, let $\sqcap = \wedge$, $\pi^0 = \langle 0.3, 1.0, 0.6\rangle^T$ and let $\boldsymbol{\Pi}$ be fixed at

$$\boldsymbol{\Pi} = \begin{bmatrix} 0.1 & 0.0 & 0.4 \\ 1.0 & 0.5 & 1.0 \\ 0.2 & 1.0 & 0.7 \end{bmatrix}.$$

The state vectors are:

| $t$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\vec{\pi}^t$ | 0.300 | 0.400 | 0.400 | 0.400 | 0.400 |
| | 1.000 | 0.600 | 1.000 | 0.700 | 1.000 |
| | 0.600 | 1.000 | 0.700 | 1.000 | 0.700 |
| $\mathbf{N}(\vec{\pi}^t)$ | 0.775 | 0.834 | 0.934 | 0.934 | 0.934 |

Here the normalizing element rotates between $i = 2, 3$.

3. Take the above example, but use $\langle \vee, \times \rangle$. Then:

| $t$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\vec{\pi}^t$ | 0.300 | 0.240 | 0.400 | 0.280 | 0.400 |
| | 1.000 | 0.600 | 1.000 | 0.700 | 1.000 |
| | 0.600 | 1.000 | 0.700 | 1.000 | 0.700 |
| $\mathbf{N}(\vec{\pi}^t)$ | 0.775 | 0.740 | 0.934 | 0.864 | 0.934 |

Again, a cycle is achieved, and now the nonspecificity initially decreases, and then cycles.

4. Finally consider the same example using $\langle \vee, \sqcap_m \rangle$, recalling that $x \sqcap_m y := (x + y - 1) \vee 0$ from Table 2.1. Then:

| $t$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\vec{\pi}^t$ | 0.300 | 0.000 | 0.400 | 0.100 | 0.400 |
| | 1.000 | 0.600 | 1.000 | 0.700 | 1.000 |
| | 0.600 | 1.000 | 0.700 | 1.000 | 0.700 |
| $\mathbf{N}(\vec{\pi}^t)$ | 0.775 | 0.600 | 0.934 | 0.758 | 0.934 |

### 5.5.2 Relation to Other Classes of Processes

The position of possibilistic process among the classes of processes has been illustrated in Fig. 5.1. The following general points are noted.

- Stochastic processes are direct generalizations of deterministic processes when the value set $V$ is generalized from $\{0,1\}$, the range of the characteristic function $\chi$, to the unit interval $[0,1]$, the range of the fuzzy membership $\mu$. However, these are both necessarily normal, additive processes.

- Fuzzy processes are similarly direct generalizations of nondeterministic processes moving from $V = \{0,1\}$ to $V = [0,1]$.

- Possibilistic processes result when normalization is imposed on fuzzy processes, while normalization is required for both stochastic and deterministic processes. Since for nondeterministic processes

$$\forall i, j \quad \pi_i^t, \boldsymbol{\Pi}_{i,j} \in \{0,1\},$$

  therefore all nondeterministic processes are also possibilistic processes, and are in fact **crisp possibilistic** processes. So possibility theory has the potential to provide insights about the operation of nondeterministic, that is general binary, systems. For example, Kanerva [136] uses such bit-string based systems to explore the high dimensionality, low cardinality, properties of such systems as genetic algorithms, cellular automata, and associative memories.

- Thus possibilistic processes proper (that is, non-crisp possibilistic processes) are also direct generalizations of nondeterministic processes, in effect, gradually weighted nondeterministic processes. And it is significant to note that stochastic processes are *not* generalizations of nondeterministic processes, despite years of commitment to them as the sole embodiment of machines with quantified uncertainty.

Finally, note that not all fuzzy relations which are normal in the sense of fuzzy sets are normal in the sense of possibilistic transition matrices (while the converse *is* true). Treating $\boldsymbol{\Pi}$ as a fuzzy relation, fuzzy set normalization (2.41) only requires that

$$\bigvee_{\omega_i, \omega_j \in \Omega} \pi(\omega_i | \omega_j) = 1,$$

which is that there is *some* unitary element in $\boldsymbol{\Pi}$, not that there is some unitary element in *each column* of $\boldsymbol{\Pi}$.

#### 5.5.2.1 Possibilistic Processes and Fuzzy Theory

Whereas we have emphasized in Sec. 2.9 that probability distributions are fuzzy sets as possibility distributions are, and therefore that possibility theory has no privileged relation to fuzzy theory, this appears to be counteracted here in that possibilistic processes are a case of fuzzy processes, but stochastic processes are not. I believe that this is the result of yet another unfortunate historical accident, which is the conflation of $\langle \vee, \wedge \rangle$ as a dual conorm/norm pair, with $\langle \vee, \wedge \rangle$ as a semiring on $[0, 1]$.

First, as argued in Sec. 2.9.5.5, one reason why possibility theory has been conflated with fuzzy theory is the fact that while $\vee$ is naturally the *only* possibilistic distribution operator, it is *also* the primary fuzzy union operator. $\wedge$ is the dual norm of $\vee$, so naturally many fuzzy applications have focused on systems which use $\langle \vee, \wedge \rangle$ pairs.

But $\langle \vee, \wedge \rangle$ is also a semiring. In fact by (5.7), it is the *only* conorm semiring where $\sqcup$ and $\sqcap$ are *also* dual. But in process theory (unlike, say, fuzzy logic) the DeMorgan property of dual $\langle \sqcup, \sqcap \rangle$ pairs is not utilized. In fact, negation itself has not been used so far in process theory! Rather it is the distributivity of semirings which is crucial, as shown in many of the proofs in this chapter.

Thus it is not possibility theory which has a special status with respect to fuzzy theory, but rather $\langle \vee, \wedge \rangle$ which has a special status as the only pair of operators which *both* form a conorm semiring *and* a dual conorm/norm pair. It is easy to overlook that possibilistic processes can be based on *any* $\langle \vee, \sqcap \rangle$ semiring, as long as it is normal.

Similarly, it is easy to overlook the fact that fuzzy theory can use union operators other than $\vee$, and that fuzzy relation composition (possibilistic or otherwise) can use semirings other than $\langle \vee, \sqcap \rangle$, let alone $\langle \vee, \wedge \rangle$. For example, we have shown that $\langle \sqcup_m, \times \rangle$ can be used for fuzzy matrix composition, since $+$ is a conorm when restricted to $[0, 1]$ (which is *always* the case in stochastic processes). And there may be yet other conorm semirings. In fact, by this argument, ordinary matrix arithmetic should be considered as a case of fuzzy matrix composition!

So it seems that the identification of general fuzzy processes with even $\langle \vee, \sqcap \rangle$ semirings may be too restrictive: whereas possibility theory is indeed wedded to $\vee$, fuzzy theory in general is not and should not be. All conorm processes rely on conorm semirings with $V = [0, 1]$, and use the fuzzy relation $\tilde{\sigma}$ to operate on the fuzzy set $\tilde{\phi}^0$ to create new fuzzy sets $\tilde{\phi}^t$, be they possibility distributions, probability distributions, or general fuzzy sets. So from a "pure" GIT perspective, we should call $\langle \vee, \sqcap \rangle$ processes **general possibilistic**, to distinguish them from proper (normalized) possibilistic processes. As all possibilistic processes are general possibilistic, so all general possibilistic *and* all additive processes, that is all conorm processes, should be identified as "fuzzy processes".

$\langle \vee, \sqcap \rangle$ fuzzy relation composition has actually been discussed a fair amount in the literature. Kaufmann introduces it as a generalization of $\langle \vee, \wedge \rangle$ composition [138]. And Santos discusses $\langle \vee, \times \rangle$ processes in some depth [249,250]. Akiyama et al. have shown the following important result.

**Proposition 5.44** [2] Crisp general fuzzy processes are identical for all $\sqcap$.

Thus as long as $\mathcal{Z}$ is crisp (for example, for nondeterministic processes), $\langle \vee, \sqcap \rangle$ composition does not depend on the choice of $\sqcap$.

#### 5.5.2.2 Possibilistic Processes and Probability Theory

Nevertheless, the $\langle \vee, \wedge \rangle$ semirings of proper fuzzy processes are the standard form for fuzzy relation composition. $\langle \vee, \wedge \rangle$ composition is used most commonly in fuzzy applications, and is discussed in all the textbooks.

However, $\vee$ and $\wedge$ do arise in probability theory, as has been observed by Gaines and Kohout [92], among others, in that

$$A \subseteq B \quad \rightarrow \quad \Pr(A \cup B) = \Pr(A) \vee \Pr(B), \quad \Pr(A \cap B) = \Pr(A) \wedge \Pr(B).$$

This would seem to justify the idea that $\langle \vee, \wedge \rangle$ plays a strictly analogous role to $\langle +, \times \rangle$ in probability theory. There is a corresponding desire to construct possibility theory and possibilistic applications by simply taking stochastic formula and globally replacing $+$ with $\vee$ and $\times$ with $\wedge$.

While we have justified this approach for $\vee$, the same is not the case for $\wedge$. Most theoreticians are aware of this (see Cumani [41], for example), although there are some exceptions.

- Hisdal is quick to say

  > In the theory of possibility the $\wedge$ (min) operation replaces the multiplication operation of the theory of probability. [118]

- Sudkamp [275] has noted that Dempster's rule (2.61) applied to consonant random sets results in a multiplicative combination rule for possibility distributions, denoted $\pi^1 \odot \pi^2 = \pi^1 \times \pi^2$. He goes on, then, to introduce a norm-dual probabilistic sum operator $\pi^1 \star \pi^2$, calling this a "plausibilistic" approach. He then moves to define a pure "possibilistic" approach where similar dual operators are $\pi^1 \wedge \pi^2$ and $\pi^1 \vee \pi^2$, despite the fact that (at least) $\wedge$ should be generalized to $\sqcap$.

- And more popular, application-oriented texts, for example Dougherty and Giardina [50], are quick to ignore anything other than $\langle \vee, \wedge \rangle$ composition.

| | $\Pr(A \cup B)$ | $\Pr(A \cap B)$ |
|---|---|---|
| General case | $a + b - \Pr(A \cap B)$ | $\Pr(A \cap B)$ |
| $A \subseteq B$ | $a + b - (a \wedge b) = a \vee b$ | $a \wedge b$ |
| $A \perp B$ | $a + b$ | $0$ |
| $A$ independent from $B$ | $a \star b = a + b - ab$ | $a \times b$ |

Table 5.2: Structural conditions in probability, where $a = \Pr(A), b = \Pr(B)$.

But consider a more complete examination of the different values of $\Pr(A \cup B)$ and $\Pr(A \cap B)$ based on the topological relations between $A, B \subseteq \Omega$, as detailed in Table 5.2, where $a = \Pr(A)$ and $b = \Pr(B)$. While $\Pr(A \cup B)$ is commonly a conorm function of $a$ and $b$, the corresponding function for $\Pr(A \cap B)$ is almost never the dual norm. $+$ is only achieved for $A \perp B$, for which $\Pr(A \cap B) = 0$, not $\Pr(A) \times \Pr(B)$. $\times$ is achieved only for stochastic independence, which results in $\star$ for the conorm operator, not $+$.

#### 5.5.2.3 Compatibility with Stochastic Processes

The PPC principle (3.5) is consistent with the operation of possibilistic processes. In particular, a stochastic process which is then converted to a possibilistic process according to GK-compatibility is equivalent to one which is converted first, and then operated by possibilistic formulae.

**Theorem 5.45 (Stochastic-Possibilistic Process Compatibility)** Let $\mathcal{R} = \langle \vee, \sqcap \rangle$, and assume fuzzy matrices $R_{n\times n} = [R_{ik}], S_{n\times n} = [S_{kj}]$ (no normalization of any kind assumed). Let $\overline{R} := [\,\overline{R_{ik}}\,]$, where

$$\overline{R_{ik}} := \begin{cases} 1, & R_{ik} > 0 \\ 0, & R_{ik} = 0 \end{cases},$$

according to PPC (and similarly for $\overline{S}$), and let $R \cdot S$ denote matrix composition under an additive semiring (i.e., standard matrix multiplication). Then $\overline{R \cdot S} = \overline{R} \circ \overline{S}$.

**Proof:** Let $T_{n\times n} = [T_{ij}] := R \cdot S$, so that $\overline{R \cdot S} = \overline{T} = [\,\overline{T_{ij}}\,]$. Also let $U = [U_{ij}] := \overline{R} \circ \overline{S}$. We need to show that $\forall i, j, \overline{T_{ij}} = U_{ij}$. First, $T_{ij} = \sum_{k=1}^{n} R_{ik} \times S_{kj}$, so that

$$\overline{T_{ij}} = \begin{cases} 0, & \underset{1 \le k \le m}{\forall} \quad R_{ik} = 0 \text{ or } S_{kj} = 0 \\ 1, & \text{Otherwise} \end{cases}.$$

Then

$$U_{ij} \quad = \quad \bigvee_{k=1}^{m} \left( \overline{R_{ik}} \wedge \overline{S_{ik}} \right) = \begin{cases} 0, & \underset{1 \le k \le m}{\forall} \overline{R_{ik}} = 0 \text{ or } \overline{S_{ik}} = 0 \\ 1, & \text{Otherwise} \end{cases}$$

$$= \begin{cases} 0, & \forall_{1 \leq k \leq m} R_{ik} = 0 \text{ or } S_{ik} = 0 \\ 1, & \text{Otherwise} \end{cases}$$
$$= \overline{T_{ij}}.$$

■

### 5.5.3 Conditional Possibility

The concept of conditional possibility was introduced in Sec. 5.3.2.4 in the definition of possibilistic processes (5.35). Fuller consideration to this idea will be given here.

Although possibility theory is usually seen as an extension of fuzzy theory, some researchers have considered it in its information theoretic form, and therefore tried to develop concepts, like conditionality, that are analogous to similar ideas in probability theory. These ideas were first introduced in 1978 in a series of papers beginning with Zadeh's founding paper for possibility theory [325], and continued by Hisdal [118] and Nguyen [191].

As above, we generalize this discussion slightly to discuss concepts of jointness, marginality, and conditionality in the context of general fuzzy measures with distributions, and operations on semirings. Then the specific concepts of conditional possibility are introduced.

In this section assume two finite universes $\Omega_1 := \{x_i\}, \Omega_2 := \{y_j\}$ with $A \subseteq \Omega_1, B \subseteq \Omega_2$. Let $\nu$ be a fuzzy measure with distribution $q$, and let $\mathcal{R} := \langle \sqcup, \sqcap \rangle$ be a conorm semiring where $\sqcup$ is the distribution operator of $q$.

**Definition 5.46 (Joints)** A **joint fuzzy measure** $\nu: 2^{\Omega_1 \times \Omega_2} \mapsto [0,1]$ is a fuzzy measure on $\Omega_1 \times \Omega_2$. Then the **joint distribution** $q: \Omega_1 \times \Omega_2 \mapsto [0,1]$ is

$$q(x_i, y_j) := \nu(\{\langle x_i, y_j \rangle\}).$$

**Definition 5.47 (Marginals)** Given a joint fuzzy measure $\nu$, then $\nu^1: 2^{\Omega_1} \mapsto [0,1]$ is a **marginal fuzzy measure** and $q^1: \Omega_1 \mapsto [0,1]$ is a **marginal distribution** where

$$\begin{aligned} \nu^1(A) &:= \nu(A \times \Omega_2) = \bigsqcup_{\langle x_i, y_j \rangle \in A \times \Omega_2} q(x_i, y_j), \\ q^1(x_i) &:= \bigsqcup_{y_j} q(x_i, y_j) = \nu(\{x_i\} \times \Omega_2). \end{aligned} \tag{5.48}$$

and similarly for $\nu^2$ and $q^2$.

**Definition 5.49 (Conditionals)** Assume a joint fuzzy measure $\nu$ and distribution $q$. Then the **conditional fuzzy measures**

$$\nu(\cdot|B)\colon 2^{\Omega_1} \mapsto [0,1], \qquad \nu(\cdot|A)\colon 2^{\Omega_2} \mapsto [0,1]$$

and **conditional distributions**

$$q(\cdot|y_j)\colon \Omega_1 \mapsto [0,1], \qquad q(\cdot|x_i)\colon \Omega_2 \mapsto [0,1]$$

are defined by the functional equations

$$\begin{aligned} \nu(A,B) &= \nu^1(A) \sqcap \nu(B|A) = \nu^2(B) \sqcap \nu(A|B), \\ q(x_i,y_j) &= q^1(x_i) \sqcap q(y_j|x_i) = q^2(y_j) \sqcap q(x_i|y_j). \end{aligned} \tag{5.50}$$

The following corollary has already been used in its possibilistic form in (5.36) in the definition of possibilistic processes.

**Corollary 5.51 (Conditional Normalization)**

$$\forall y_j, \quad \bigsqcup_{x_i} q(x_i|y_j) = 1, \qquad \forall x_i, \quad \bigsqcup_{y_j} q(y_j|x_i) = 1.$$

**Proof:** From the definition of marginals (5.48), $\bigsqcup_{y_j} q(x_i,y_j) = q^1(x_i)$. Similarly, from the definition of conditional distributions (5.50) and the distributivity of $\sqcap$ over $\sqcup$,

$$\bigsqcup_{y_j} q(x_i,y_j) = \bigsqcup_{y_j} q^1(x_i) \sqcap q(y_j|x_i) = q^1(x_i) \sqcap \left( \bigsqcup_{y_j} q(y_j|x_i) \right).$$

Therefore $\bigsqcup_{y_j} q(y_j|x_i) = 1$ because 1 is the identity for $\sqcap$. The analogous argument holds for the other case. ■

**Corollary 5.52**

$$q(x_i|y_j) \geq q(x_i,y_j) \leq q(y_j|x_i)$$

**Proof:** In general for $x, y \in \mathbb{R}$,

$$x \sqcap y \leq x \wedge y, \qquad x \wedge y \leq x, \qquad x \wedge y \leq y.$$

Therefore

$$q(x_i,y_j) = q^1(x_i) \sqcap q(y_j|x_i) \leq q^1(x_i) \wedge q(y_j|x_i) \leq q(y_j|x_i).$$

A similar argument holds for $q(x_i,y_j) \leq q(x_i|y_j)$. ■

**Proposition 5.53 (Conditional Probability)** When $\nu = \Pr, q = p$, and $\langle \sqcup, \sqcap \rangle = \langle +, \times \rangle$, then

$$\begin{aligned} p(x_i, y_j) &= \Pr(\{\langle x_i, y_j \rangle\}), \\ \Pr^1(A) &= \Pr(A \times \Omega_2) = \sum_{\langle x_i, y_j \rangle \in A \times \Omega_2} p(x_i, y_j), \\ p^1(x_i) &= \sum_{y_j} p(x_i, y_j) = \Pr(\{x_i\} \times \Omega_2), \\ \Pr(A, B) &= \Pr^1(A) \times \Pr(B|A) = \Pr^2(B) \times \Pr(A|B), \\ p(x_i, y_j) &= p^1(x_i) \times p(y_j|x_i) = p^2(y_j) \times p(x_i|y_j). \end{aligned}$$

**Proposition 5.54 (Conditional Possibility)** When $\nu = \Pi, q = \pi$, and $\langle \sqcup, \sqcap \rangle = \langle \vee, \sqcap \rangle$, then $\sqcap$ remains a free parameter, so that

- [155] Given a **joint possibility measure** $\Pi \colon 2^{\Omega_1 \times \Omega_2} \mapsto [0,1]$ on $\Omega_1 \times \Omega_2$, then the **joint possibility distribution** $\pi \colon \Omega_1 \times \Omega_2 \mapsto [0,1]$ is

$$\pi(x_i, y_j) := \Pi(\{\langle x_i, y_j \rangle\}).$$

- [155] Given a joint possibility measure $\Pi$, then $\Pi^1 \colon 2^{\Omega_1} \mapsto [0,1]$ is a **marginal possibility measure** and $\pi^1 \colon \Omega_1 \mapsto [0,1]$ is a **marginal possibility distribution** where

$$\begin{aligned} \Pi^1(A) &:= \Pi(A \times \Omega_2) = \bigvee_{\langle x_i, y_j \rangle \in A \times \Omega_2} \pi(x_i, y_j), \\ \pi^1(x_i) &:= \bigvee_{y_j} \pi(x_i, y_j) = \Pi(\{x_i\} \times \Omega_2). \end{aligned}$$

  and similarly for $\Pi^2$ and $\pi^2$.

- [67] Assume a joint possibility measure and distribution, and a norm $\sqcap$. Then the **conditional possibility measures**

$$\Pi(\cdot|_\sqcap B) \colon 2^{\Omega_1} \mapsto [0,1], \qquad \Pi(\cdot|_\sqcap A) \colon 2^{\Omega_2} \mapsto [0,1]$$

  and **conditional possibility distributions**

$$\pi(\cdot|_\sqcap y_j) \colon \Omega_1 \mapsto [0,1], \qquad \pi(\cdot|_\sqcap x_i) \colon \Omega_2 \mapsto [0,1]$$

  are defined by the functional equations (a form of "possibilistic Bayes theorem")

$$\begin{aligned} \Pi(A, B) &= \Pi^1(A) \sqcap \Pi(B|_\sqcap A) = \Pi^2(B) \sqcap \Pi(A|_\sqcap B), \\ \pi(x_i, y_j) &= \pi^1(x_i) \sqcap \pi(y_j|_\sqcap x_i) = \pi^2(y_j) \sqcap \pi(x_i|_\sqcap y_j). \end{aligned}$$

**Proposition 5.55** Some major classes of conditional possibility distributions are

$$\pi(x_i|_\wedge y_j) = \begin{cases} [\pi(x_i, y_j), 1], & \pi(x_i, y_j) = \pi^2(y_j) \\ \pi(x_i, y_j), & \pi(x_i, y_j) < \pi^2(y_j) \end{cases},$$

$$\pi(x_i|_\times y_j) = \frac{\pi(x_i, y_j)}{\pi^2(y_j)},$$

$$\pi(x_i|_{\sqcap_m} y_j) = \begin{cases} [0, 1 - \pi^2(y_j)], & \pi(x_i, y_j) = 0 \\ 1 - \pi^2(y_j) + \pi(x_i, y_j), & \pi(x_i, y_j) > 0 \end{cases},$$

and correspondingly for $\pi(y_j|_\sqcap x_i)$.

Note that conditional possibility is frequently not determinative. Cooman[2] has shown that it will only be so for norms such that

$$\forall x, y, z \in [0, 1], \quad x \sqcap y \leq x \sqcap z \rightarrow x = 0 \text{ or } y < z.$$

This is not an overwhelming limitation, however. By comparison, a unique conditional probability is not always available, either. In particular, the conditional probability $p(x|y) = p(x, y)/p(y)$ is not defined when $p(y) = 0$.

There are a variety of different approaches to this material, and still some outstanding issues. For example:

- Bouchon [18] defines joint distributions in terms of (deterministic) conditionals, rather than the other way around;
- Ramer [228] suggests a definition of conditional possibility by a method similar to the maximum plausibility focused consistent transformation normalization method presented in Sec. 2.8.1.2;
- Comman et al. [38] unify joint and conditional possibility using a measure-theoretic approach;
- There is also a strong developing movement in so-called "measure-free conditionals" or "conditional events", which views the construct $x_i|y_j$ as a (non-deterministic) event in an abstract algebra [67, 103, 105];
- From the MUP, and the fact that $\wedge \geq \sqcap$ for all norms $\sqcap$, Ramer and Puflea-Ramer [232] have justified $\wedge$ as the appropriate $\sqcap$ operator. Furthermore, **possibilistic non-interaction** is frequently defined [155] as the condition

  $$\pi(x_i, y_j) = \pi^1(x_i) \wedge \pi^2(y_j).$$

[2] Personal communication.

Some of the properties of "min-related" fuzzy numbers have been explored by Rao and Rashed [233].

- Dubois and Prade [67] have suggested that

$$\pi(x_i, y_j) = \pi(y_j) \rightarrow \pi(x_i |_\wedge y_j) := 1$$

on the basis of applying the MUP to $[\pi(x_i, y_j), 1]$;

- And Zadeh [325] originally suggested

$$\pi(x_i | y_j) := \pi(x_i, y_j).$$

Aside from the generalization to general fuzzy measures, distributions, and conorm semirings made here, we have little to add to the specific issues surrounding the nature of conditional possibility. Our interest is in the use of conditional possibility in possibilistic processes, their meaning and measurement. Recall that $\boldsymbol{\Pi}$ is a vector of conditional possibility distributions on $\Omega^2$, each of which is normalized so that

$$\forall \omega_j \in \Omega, \quad \bigvee_{\omega_i \in Q} \pi(\omega_i | \omega_j) = 1.$$

So in the context of possibilistic processes, this condition states that no matter what the current state, there must be at least one state to which it is completely possible to transit, one out arrow weighted 1 in a possibilistic state transition diagram.

Were this not to hold, then formally $\boldsymbol{\Pi}$ would simply be a general fuzzy relation, with no possibilistic constraints whatsoever. And semantically, a process might arrive in a state $\omega_j$ for which $\not\exists i, \boldsymbol{\Pi}_{ij} = 1$, so that the process would be "possibilistically blocked" from any future progress. Note that this is different from an "absorbing" state, which is indefinitely transiting only to itself. So when transition possibilities are determined from measurement, normalization is expressed as the requirement that some state be identified for future progress, or in other words as a commitment that future measurements can indeed be made.

### 5.5.4 Other Approaches

Finally, some of the approaches that other researchers have taken to some of these issues should be mentioned.

- Gaines and Kohout [92] discussed both possibilistic state and transition normalization, but they did so only in the context of general automata theory on semirings, and not in the special context of possibility theory, possibility distributions, and conditional possibility distributions. This is despite the fact that in the same paper they also laid out many of the key *semantic* conditions for possibility theory which were used so heavily in Chap. 3.

- Santos [246] identified $\mathcal{Z}_\pi$ as a **restricted fuzzy process**, but gave it no further consideration.

- Yager [317] generalizes the fuzzy integral, which, like fuzzy relation composition, was also originally based on $\langle \vee, \wedge \rangle$ composition, to general $\langle \sqcup, \sqcap \rangle$ pairs, but is not concerned with distributivity. This is contrasted with the approach of Cooman et. al [38], who generalize fuzzy integrals to "possibilistic integrals" using conorm semirings on lattices.

- Dal Cin [42,43] discusses fuzzy automata, but these are of a decidedly different formal nature from the systems discussed in this chapter.

- Wee and Fu [302] use the term "normalized fuzzy automata" to refer to a fuzzy automata which is *stochastically* transition normal.

## 5.6 Possibilistic Systems

As with stochastic processes, possibilistic processes form the mathematical core around which other possibilistic systems and machines can be developed.

### 5.6.1 Automata

Automata have been mentioned many times in the above sections, but will be mathematically defined here by extending general processes to include input and/or output functions.

**Definition 5.56 (Automaton)** Assume a finite **input alphabet** $Y := \{y_k\}$ and define the transition function as $\sigma \colon \Omega \times Y \times \Omega \mapsto [0,1]$. For a given $y_k \in Y$, then denote $\sigma(\cdot, y_k, \cdot) \colon \Omega_2 \mapsto [0,1]$ as the projection of $\sigma$ through $y_k$. Then the system $\mathcal{A} := \langle \Omega, Y, \oplus, \otimes, \sigma, \phi^0 \rangle$ is an **automaton** if $\forall y_k \in Y, \langle \Omega, \oplus, \otimes, \sigma(\cdot, y_k, \cdot), \phi^0 \rangle$ is a process.

An automaton is a process whose state transitions are paramaterized by its input. Each input symbol $y_k$ establishes a different transition function $\sigma(\cdot, y_k, \cdot)$.

Some of the other familiar forms of automata with output [273] are also available.

**Definition 5.57 (Mealy Automaton)** Assume an automaton $\mathcal{A}$, a finite **output alphabet** $Z := \{z\}$, and define an **output function** $\lambda: \Omega \times Y \times Z \mapsto [0,1]$. Then $\langle \Omega, Y, Z, \oplus, \otimes, \sigma, \phi^0, \lambda \rangle$ is a **Mealy automaton**.

**Definition 5.58 (Moore Automaton)** A Mealy automaton $\mathcal{A}$ is a **Moore automaton** if

$$\forall_{y_{k_1}, y_{k_2} \in Y} \forall_{\omega_i \in \Omega} \forall_{z \in Z} \lambda(\omega_i, y_{k_1}, z) = \lambda(\omega_i, y_{k_2}, z).$$

$\lambda(\omega_i, y_k, z)$ represents the degree to which $z$ is output given being in state $\omega_i$ and receiving input symbol $y_k$.

**Possibilistic automata** follow naturally when $\langle \oplus, \otimes \rangle = \langle \vee, \sqcap \rangle$ and $\phi^t$ and $\sigma$ are normal and transition normal respectively.

General automata have actually been discussed very little in the literature, which concentrates almost entirely on the special cases discussed in Sec. 5.3. Exceptions are Santos [251], who early on explored areas beyond $\langle \vee, \wedge \rangle$ composition, and Gaines and Kohout [92], which again forms the inspiration for much of this work.

### 5.6.2 Possibilistic Monte Carlo Methods

Another major component of stochastic process theory is the ability to move from the distribution-level meta-state description $\vec{p}^{\,t}$ to a specific state-level description. That is, even though the state vector $\vec{p}^{\,t}$ is available at each time $t$, it is desirable to select a *single* state $\omega_i \in \Omega$ as being selected at time $t$. This is done by **Monte Carlo methods**, which through the generation of random numbers, simulate the movement of specific states forward in time in accordance with the probability distributions in question.

First we state the definition of the standard (probabilistic) Monte Carlo method.

**Definition 5.59 (Monte Carlo Method)** Assume a cumulative probability distribution function $P: \mathbb{R} \mapsto [0,1]$, where

$$P(x) := \Pr([-\infty, x]) = \int_{-\infty}^{x} p(y)\, dy$$

for some probability distribution $p: \mathbb{R} \mapsto [0,1]$, and let $U$ be a uniform random variable on $[0,1]$. Then the **Monte Carlo method** selects that $x_0 \in \mathbb{R}$ such that $\Pr([-\infty, x_0]) = U$, which is $P^{-1}(U)$.

This definition is dependent on the fact that $P$ is both a probability measure and a monotonic nondecreasing transform of $p$. $P$, being the definite integral of a non-negative probability distribution, is bijective on $[0,1]$. Thus $P^{-1}(U)$ exists for a unique $x_0$, so that $P^{-1}$ maps a probability $U \in [0,1]$ to a unique event $[-\infty, x_0]$.

When making the movement to the possibilistic case, the situation is not so straightforward. There are a variety of approaches, each with advantages and disadvantages.

First, note that in the probabilistic case, $P$ being bijective depends on the ordering $\leq$ of $\mathbb{R}$ and the topological structure of the sets $[-\infty, x]$, so that

$$x_1 \leq x_2 \to [-\infty, x_1] \subseteq [-\infty, x_2].$$

Thus the sets $[-\infty, x]$ form a nest, and $P$ actually has a form similar to a possibility distribution. Where $P$ is the sum (integral, in the continuous case) of probabilities $p$ on $x \in \mathbb{R}$, $\pi$ is the sum of "probabilities" (evidence values) $m$ on $A_j \in 2^\Omega$.

It therefore might make sense to have a possibilistic Monte Carlo method select $\omega_0 \in \Omega$ such that $\omega_0 = \pi^{-1}(U)$. However, this overlooks the very different semantics of possibility and probability, as has been discussed repeatedly. $U$ is a probability, and it makes little sense to compare it to a possibility value.

Moreover, while generally

$$x_1 \neq x_2 \to P(x_1) \neq P(x_2)$$

(this fails to hold only when $\forall x \in [x_1, x_2], p(x) = 0$), from (2.96) the corresponding condition for possibility distributions

$$x_1 \neq x_2 \to \pi(x_1) \neq \pi(x_2)$$

only holds when $\pi$ is complete. Otherwise, $\pi^{-1}(U)$ may not be unique. Complete possibility distributions are generally rare, requiring (at least) a point focus for a core. Note that this is usually not the case for possibilistic histograms.

In fact, it is a general principle in possibility theory that results tend to be interval-valued. Further, in keeping with the random set interpretation of possibility theory, it is perhaps better to concentrate on a Monte-Carlo method at the level of $\mathcal{S}$. In this approach, as suggested by Chanas and Nowakowski [31], $U$ would be used only to select an entire *focal element*.

**Definition 5.60 (Random Set Monte Carlo Method)** [31] Given a random set $\mathcal{S}$, consider the $m_j$ as probabilities, and then:

1. Select an $A_j$ by a probabilistic Monte Carlo method;

2. Select $\omega_0 \in A_j$ by iterating a Monte Carlo method from a uniform MEP distribution on $A_j$.

**Proposition 5.61** [31] $\Pr(\omega_0 = \omega) = p^{\mathcal{S}}(\omega)$, recalling that $p^{\mathcal{S}}$ is the maximum entropy probability distribution on $\mathcal{S}$ (2.122).

**Proposition 5.62** [31] Consider a possibility distribution $\pi$, with a consonant constructed random set $\mathcal{S}^\pi$ from (2.128), as a fuzzy set $\tilde{\pi}$. Then selecting $\omega_0$ from the random set Monte Carlo method (5.60) is equivalent to:

1. Select the alpha cut $\tilde{\pi}_U$;
2. Select $\omega_0$ by a Monte Carlo method from a uniform MEP distribution on $\tilde{\pi}_U$.

**Proposition 5.63** Selecting $\omega_0$ from (5.62) is equivalent to selecting $\omega_0$ by a Monte Carlo method from a uniform MEP distribution on $\mathcal{S}(U)$, as determined by (2.134).

Wong and Shen [308] suggest that $\omega_0$ should be chosen from a general fuzzy set as that state which has maximal membership grade. In the context of possibility distributions, this is equivalent to selecting $\omega_0$ by a Monte Carlo method from a uniform MEP distribution on $\mathbf{C}(\pi)$. The disadvantage of this method is that no information about properly possible values $0 < \pi(\omega) < 1$ is used. However, this method is to be chosen by GK-compatibility. Under PPC, it is equivalent to selecting $\omega_0$ by a Monte Carlo method from a uniform MEP distribution on $\{\omega : p(\omega) > 0\}$.

An alternative would be to select $\omega_0$ by a Monte Carlo method from a uniform MEP distribution on $\mathbf{U}(\pi)$. But as above, this would also ignore all information about the possibilistic structure of $\pi$.

### 5.6.3 Possibilistic Markov Processes

Note that in a possibilistic process a state vector $\vec{\pi}^t$ is a function only of the time-invariant transition matrix $\boldsymbol{\Pi}$ and the previous time state vector $\vec{\pi}^{t-1}$. It is interesting to consider cases where more past states affect the current state, for example by considering $\boldsymbol{\Pi}$ as a vector of conditional possibility distributions of the form $\pi(\omega_i^t | \omega_j^{t-1}, \omega_k^{t-2})$. Analogously to stochastic processes, the above example can be considered as a **possibilistic Markov process** of degree 2. Thus $\mathcal{Z}_\pi$ is a possibilistic Markov process of degree 1, and generalizations to degree $n$ are evident.

### 5.6.4 Other Forms

As with stochastic processes, a variety of other forms of possibilistic systems have either been considered in the literature already (although not in this context of mathematical possibilistic process theory), or suggest themselves immediately.

- Notice that the transition graph of a possibilistic process, as in Fig. 5.2, has the form of a network diagram. Possibilistic processes can therefore be generally described as **possibilistic networks**, where states indicate nodes of the network and transition possibilities are non-additive, possibilistically normal, connection weights. Since these weights are conditional possibilities, possibilistic networks are possibilistic correlates to the Bayesian networks of classical information theory.

- Dubois and Prade [73] have considered possibilistic hypergraphs.

- Yager [316] has introduced possibilistic production systems.

- While coin toss and ball-in-urn thought experiments are a staple of probability theory, there are as of yet no such simple, canonical examples for possibility theory. However, it is interesting to consider those classical examples converted to possibility theory using the UIP. Klir has done this [153] for a simple ball-in-urn problem.

- Possibilistic analogs to cellular automata [37,287], L-systems [177], and neural nets [243] should be considered. Fuzzy Petri nets [208] have already been considered by Looney [178] and Person [207].

# Chapter 6

# Implementation of Possibilistic Models

You'll hear that programmers succeeded in bringing the Central Computer under control, cutting its higher reasoning centers while new programs could be written, leaving the merely mechanical parts of me intact so I could continue running things. They probably believe that, too, but they're wrong. If their schemes had reached fruition, I wouldn't be talking to you now because we'd both be dead, and so would every other human soul on Luna.

— *John Varley*

Mathematics has traditionally been concerned with continuous systems, for example functions on $\mathbb{R}^n$. So while both mathematics and systems theory are concerned with systems of all types, both discrete and continuous, systems theory has tended as a matter of course to focus more on discrete systems.

The distinction between discrete and continuous systems is closely mirrored by that between digital and analog machines. So it is not surprising that there is an intimate relationship between systems theory and computer technologies, at many levels and in many ways.

First, there is a strong historical relationship. Many of the first systems theorists and cyberneticians, such as von Neumann and Weiner, were also some of the pioneering computer scientists. Also, cybernetics was specifically intended to develop the mind-machine analogy, in particular the metaphor of the brain as a digital computer.

But beyond that, digital technology is both appropriate and even necessary for research on the kinds of discrete systems typically studied by systems science. Formally, computer systems (as finite Turing machines) are in fact cases of discrete systems. This class also includes the kinds of systems with high dimensionality considered by Kampis (components systems) [130] and Kanerva (the Hamming metric space of bit strings) [136].

Moreover, in virtue of their high combinatorial complexity, discrete systems do not generally yield to the kinds of simple mechanical or electrical analog models that continuous systems, such as systems of differential equations, do. In fact, their physical representation or modeling *requires* digital technology.

Klir has observed that systems methods cannot be fully developed or explored without computer-based implementations of them.

> Systems knowledge can also be obtained experimentally. Although systems (knowledge structures) are not objects of reality, they can be simulated on computers and in this sense made real. We can then experiment with the simulated systems for the purpose of discovering or validating various hypotheses in the same way as other scientists do with objects of their interest in their laboratories. In this sense, computers, may be viewed as laboratories of systems science. Experimentation with systems on computers is not merely possible, but it may give us knowledge that is otherwise unobtainable. [149, p. 102]

He has also suggested that the lack of computing machines has (until recently) allowed *only* for the full development of either continuous or very simple discrete mathematical systems. For example, he remarks[1] that Laplace had considered non-additive "probabilities" (in other words the kinds of belief and plausibility values later developed by Dempster and Shafer into evidence theory), but was not able to continue because of the enormous combinatorial complexity of the calculations on $2^{\Omega}$ that would have been required.

As described by Horgan [123], as the complexity of problems increases, this situation is becoming common generally in mathematics. The result is the growth of so-called "experimental" or "computer-aided" mathematics, where computer-based tools are used to empirically investigate the properties of mathematical systems.

Therefore, in this chapter an architecture for the computer-based implementations of possibilistic models in an object-oriented environment, in particular the C++ programming language, is proposed. Such a system is crucial not only as a platform for the application of possibilistic qualitative modeling (discussed below in Chap. 7), but also for the empirical investigation of the properties of possibilistic

[1] Personal communication.

processes. Examples of many of the open questions that require empirical investigation include the effects of the choices of continuous approximations and focus values in the possibilistic measurement methods from Chap. 4, and the evolution of the values of information measures under the operation of possibilistic processes and normalization methods.

## 6.1 Computer-Aided Systems Theory (CAST)

In presenting this architecture, we will also discuss some of the issues related to the development of computer-based modeling environments and implementations of systems theoretic methods in general. This task is perhaps best exemplified by the Computer-Aided Systems Theory (CAST) school. CAST includes the work of (especially) Pichler and his colleagues [184, 210–214, 219, 220], Zeigler and his colleagues in their approach to discrete event modeling [218,326,327], Paul Fishwick [85–87], Tuncer Ören [195, 196], and others [281].

While CAST can be broadly characterized as any attempt at computer-based implementations of systems theoretic methods, it has tended to focus on the construction of generalized, integrated modeling environments that encompass multiple methodologies, allowing for their representation in a common framework and for transformations among them.

In a well-engineered CAST environment, implementations of specific, complex systems theoretic methods would be built from and depend upon those of more general, foundational methods. In fact, a general purpose systems theoretic modeling language (built around such concepts as state spaces, relations, functions, and deterministic processes, for example) could provide a robust environment in which to implement more detailed methods (for example Turing machines or Petri nets). This is the approach of Pichler's group, for example, where basic implementations of finite state machines are then used to develop a variety of specialized cases.

Existing CAST implementations (for example Pichler's and Zeigler's) are deterministic. The extension of these implementations, and the development of new environments, to include representations of indeterminism, uncertainty, and information is crucial. For example, existing systems could not implement neural networks with (stochastic) noise.

It is also clear that the fundamental categories for the representation of uncertainty and information should be included in that first category of basic systems methods, on the basis of which more complex methods are built. These categories should include the entire repertoire of GIT, allowing, for example, the use of methods from probability and statistics and fuzzy theory, as well as random sets and possibility theory.

GIT-based CAST implementations should allow the handling of hybrid sources and representations of uncertainty, the integration of multiple sources of information, and the transformation between representational forms of information, as with the UIP. For example, Klir's General Systems Problem Solver (GSPS) [28,81,145,194] was designed specifically to accommodate both probabilistic and possibilistic representations of information, and is best implemented in a GIT-based general systems theoretic CAST environment.

There has recently been an explosion of implementations of fuzzy systems methods for both the commercial and academic markets (Janzen [126] and Sosnowski and Pedrycz [272] are examples). The same is not the case for general GIT methods, however, and certainly not in possibility theory. One exception is the work of Galway [93], who has implemented a system for manipulating random subsets of $\mathbb{R}^2$.

## 6.2 Object-Oriented Environments

One of the most successful programming paradigms of recent years is the **object-oriented** approach [36,40]. This methodology is based on the concepts of **objects**, which are complex data elements, and **classes**, or "intelligent data types" for objects, which isolate type-specific procedures within type-specific levels. Logical relations among classes allow for the **inheritance** of procedures from more general classes to their specialized cases. Classes and objects have **attributes**, either other objects (data attributes) or **methods** (procedural attributes). A class **invariant** is a logical condition which must always be true of every object of the class in order for it to be existing in a legal state.

Popular object-oriented programming languages include Smalltalk [101], Eiffel [7,39], Objective C [40], and Loops [141] and Scoops [327] (object-oriented extensions to Lisp and Scheme respectively). The target language for the proposed architecture below is C++ [21,274]. It was selected for its popularity and efficiency, and because of the availability of standard, inexpensive compilation environments and software support libraries.

The main results of this chapter are summarized in Figs. 6.1–6.3, which show Entity-Relationsip (ER) diagrams of the proposed class hierarchies. ER diagrams [36,163] are a common form for the representation of the most prominent semantic relations among classes. The ER diagrams used here are a slight modification of the standard form presented by Coad and Yourdon [36]. Each node denotes a class (data type). Nodes are linked by labeled arcs, each indicating one of the following relations, where $X$ and $Y$ are classes:

- $X \stackrel{\text{is-a}}{\longrightarrow} Y$: $Y$ inherits from $X$, so that $Y$ is a **specification** of $X$ and has all the properties of $X$. $X$ is called a **parent** and $Y$ a **child**. For example

  $$\texttt{bird} \stackrel{\text{is-a}}{\longrightarrow} \texttt{robin}$$

  would require that `robins` inherit from `birds`. **Multiple inheritance** results when a class inherits from multiple parent classes. For engineering reasons or to capture efficiencies present in the special cases, some attributes may be implemented redundantly in child classes. For example, the formula for $\mathbf{N}(\mathcal{S})$ (2.99) is greatly simplified in the consonant case by $\mathbf{N}(\pi)$ (2.110).

- $X \stackrel{\text{collection}}{\longrightarrow} Y$: $Y$ objects are implemented as a **collection** (for example, a list, set, bag, or vector) of $X$ objects. For example

  $$\texttt{bird} \stackrel{\text{collection}}{\longrightarrow} \texttt{flock}$$

  could mean that a `flock` object is a set of `bird` objects.

- $X \stackrel{\text{has-a}}{\longrightarrow} Y$: $Y$ is a **component** of $X$, so that each $X$-object contains a $Y$ sub-object. For example

  $$\texttt{bird} \stackrel{\text{has-a}}{\longrightarrow} \texttt{wing}$$

  would require that each bird object contain a wing object. When $Y$ is a sub-object of $X$, then $Y$ may have access to $X$-specific information. For engineering or efficiency reasons, it may be that $Y$ can be implemented separately from $X$, or a $Y$ may be constructed from an $X$, copying the appropriate $X$-specific information into $Y$. In fact, at the strictly logical level this relation simply requires that each $X$ object **determines** a unique $Y$ object, or that a procedure exists to construct a $Y$ object from an $X$ object. This is the sense that will frequently be used below.

Note that while arrows always move from the more general to the more specific, this is not always mirrored by the English translation of the arcs. For example, $X \stackrel{\text{is-a}}{\longrightarrow} Y$ is read that "$Y$ is an $X$", while $X \stackrel{\text{has-a}}{\longrightarrow} Y$ is read that "$X$ has a $Y$".

## 6.3 Fundamental Classes

Each of the ER diagrams below describes a different portion of the overall architecture, and is accompanied by a set of descriptions of the classes included in the figure. Only the most basic methods and procedures are included here; supplementary methods are described in Sec. 6.4.2.

The ER diagrams and class descriptions are written in a kind of non-C++-specific class design "pseudo-code". Only the *logical* relations among the classes are described. For example, in Fig. 6.1 it is not specified exactly how a plausibility assignment is determined from a plausibility measure. It will be presumed that if $X \xrightarrow{\text{has-a}} Y$, then $Y$ will have access to $X$-specific information.

The figures share the class `Poss_Dist`, for possibility distributions, in common.

### 6.3.1 Random Sets

Fig. 6.1 shows the class hierarchy including random sets, a class of fuzzy measures on them, and those measures' distributions. The class `Random_Set` is the most general, and therefore one of the most heavily laden, classes in the proposed architecture.

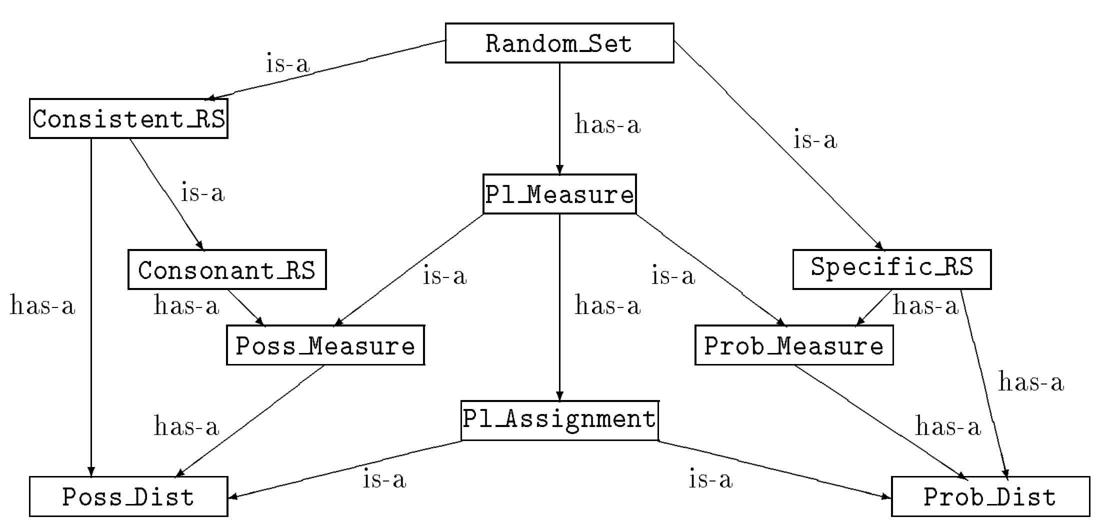


Figure 6.1: Random sets, evidence measures, and distributions.

`Random_Set` — A random set $\mathcal{S}$.

| | | |
|---|---|---|
| **Data Attributes** | `Universe` | Integer $n = \|\Omega\|$ |
| | `Card` | Integer $N = \|\mathcal{S}\|$ |
| | `Data` | A list of $\langle a_j, m_j \rangle$ pairs, with floats $m_j$ and integers $a_j$, where $a_j$ is an $n$-long bit-mask determining the characteristic function $\chi_{A_j}$ of the subset $A_j \subseteq \Omega$ |
| **Invariants** | | $1 \leq n$ |
| | | $1 \leq j \leq N \leq 2^n - 1$ |
| | | $1 \leq a_j \leq 2^n - 1$ |
| | | $0 \leq m_j \leq 1$ |
| | | $\sum_{j=1}^{N} m_j = 1$ |
| **Methods** | `Strife` | $\mathbf{S}(\mathcal{S})$ |
| | `Nonspec` | $\mathbf{N}(\mathcal{S})$ |
| | `Total` | $\mathbf{T}(\mathcal{S})$ |
| | `Core` | $\mathbf{C}(\mathcal{S})$ |
| | `Monte_Set` | Select a focal element $A_j$ |
| | `Monte` | Select a universe element $\omega_i$ |
| | `+` | The Dempster combination operator $\odot$ which combines this `Random_Set` with another, producing a combined `Random_Set` |
| | `<=` | The random set inclusion relation $\subseteq$ (2.131), a boolean reporting whether this `Random_Set` is included within another |
| | `Complete?` | Boolean: is $\mathcal{S}$ complete? |

`Consistent_RS` — A consistent random set.

| | | |
|---|---|---|
| **Invariant** | | $\mathbf{C}(\mathcal{S}) \neq \emptyset$ |

`Consonant_RS` — A consonant random set.

| | | |
|---|---|---|
| **Data Attribute** | `Ordering` | A list $\langle a_{\bar{j}} \rangle$ permuting the $a_j$ according to the inclusion relation among the $A_{\bar{j}}$ |
| **Invariants** | | $\forall 1 \leq \bar{j}_1 \leq \bar{j}_2 \leq N, A_{\bar{j}_1} \subseteq A_{\bar{j}_2}$ |
| | | $1 \leq N \leq n$ |

`Specific_RS` — A specific random set.

| | |
|---|---|
| **Invariants** | $\forall 1 \leq j \leq N, \lvert A_j \rvert = 1$ |
| | $1 \leq N \leq n$ |

`Pl_Measure` — A plausibility measure Pl on a random set.

| | | |
|---|---|---|
| **Method** | `Value` | Given $0 \leq a_j \leq 2^n - 1$, returns a float $\mathrm{Pl}^j := \mathrm{Pl}(A_j)$. |

`Poss_Measure` — A possibility measure $\Pi$.

| | |
|---|---|
| **Invariant** | $\forall a_1, a_2, \mathrm{Pl}(A_1 \cup A_2) = \mathrm{Pl}^1 \vee \mathrm{Pl}^2$. |

`Prob_Measure` — A probability measure Pr.

| | |
|---|---|
| **Invariant** | $\forall a_1, a_2, \mathrm{Pl}(A_1 \cup A_2) = \mathrm{Pl}^1 + \mathrm{Pl}^2 - \mathrm{Pl}(A_1 \cap A_2)$. |

`Pl_Assignment` — A plausibility assignment $\vec{\mathrm{Pl}}$ from a random set.

| **Data Attribute** | `Data` | A list of floats $\langle \mathrm{Pl}_i \rangle$, where $\mathrm{Pl}_i = \mathrm{Pl}(\{\omega_i\})$ |
|---|---|---|
| **Invariant** | | $1 \leq i \leq n$ |

`Poss_Dist` — A possibility distribution $\pi$.

| **Invariant** | | $\bigvee_i \mathrm{Pl}_i = 1$ |
|---|---|---|
| **Methods** | `Nonspec` | $\mathbf{N}(\pi)$ |
| | `Strife` | $\mathbf{S}(\pi)$ |
| | `Total` | $\mathbf{T}(\pi)$ |
| | `Core` | $\mathbf{C}(\pi)$ |

`Prob_Dist` — A probability distribution $p$.

| **Invariant** | | $\sum_i \mathrm{Pl}_i = 1$ |
|---|---|---|
| **Method** | `Entropy` | $\mathbf{H}(p)$ |

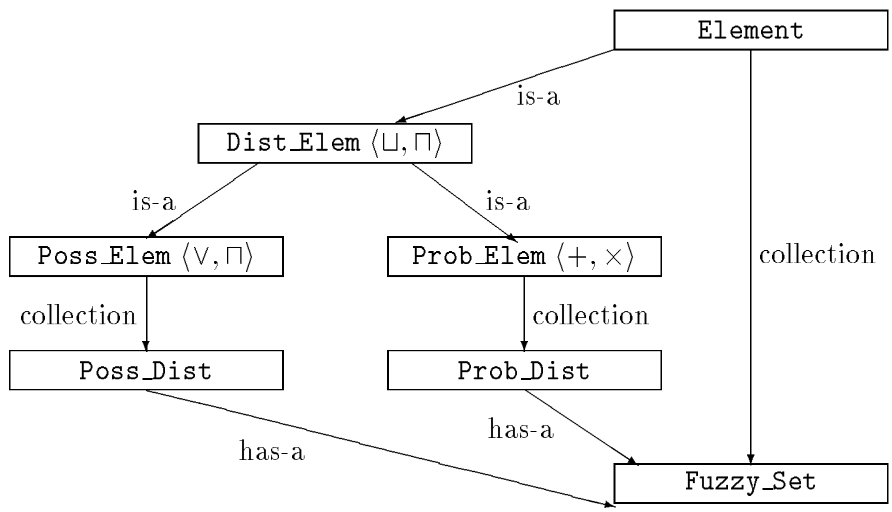


Figure 6.2: Distributions and fuzzy sets.

### 6.3.2 General Distributions

Fig. 6.2 shows the class hierarchy of distributions and fuzzy sets.

`Element` — A generic element of a distribution or fuzzy set.

| **Data Attribute** | `Value` | A floating-point "fit" (fuzzy digit), $f$. |
|---|---|---|
| **Invariant** | | $0 \leq f \leq 1$ |

`Dist_Elem` — An element of a distribution $q_i$.

| | | |
|---|---|---|
| **Methods** | `+` | The distribution operator $\oplus$ which aggregates this `Dist_Elem` with another, producing an aggregated `Dist_Elem`. At this level of generality $\oplus$ is not specified (`Dist_Elem` is a virtual base class), but it must be a conorm $\sqcup$. |
| | `*` | The combination operator $\otimes$ which combines this `Dist_Elem` with another, producing a combined `Dist_Elem`. At this level of generality $\otimes$ is not specified (`Dist_Elem` is a virtual base class), but $\langle \oplus, \otimes \rangle$ must be a conorm semiring $\langle \sqcup, \sqcap \rangle$. |

Fuzzy_Set — A collection of elements comprising $\mu$.

| | | |
|---|---|---|
| **Data Attributes** | Universe | Integer $n = \|\Omega\|$ |
| | Data | A list of Elements $\langle f_i \rangle$. |
| **Invariant** | | $1 \leq i \leq n$ |

Poss_Elem — Element of a possibility distribution $\pi_i$.

| | | |
|---|---|---|
| **Invariant** | | $f_1 \sqcup f_2 = f_1 \vee f_2$. |

Poss_Dist — A possibility distribution $\pi$.

| | | |
|---|---|---|
| **Data Attributes** | Universe | Integer $n = \|\Omega\|$ |
| | Data | A list of Poss_Elems $\langle \pi_i \rangle$ |
| **Invariant** | | $1 \leq i \leq n$ |
| | | $\bigvee_{i=1}^{n} \pi_i = 1$ |

Prob_Elem — Element of a probability distribution $p_i$.

| | | |
|---|---|---|
| **Invariant** | | $f_1 \sqcup f_2 = f_1 + f_2$. |

Prob_Dist — A probability distribution $p$.

| | | |
|---|---|---|
| **Data Attributes** | Universe | Integer $n = \|\Omega\|$ |
| | Data | A list of Prob_Elems $\langle p_i \rangle$ |
| **Invariant** | | $1 \leq i \leq n$ |
| | | $\sum_{i=1}^{n} p_i = 1$ |

Note that Poss_Dist and Prob_Dist are repeated here as collections of their elements, inheriting from Pl_Assignment from Fig. 6.1.

### 6.3.3 Possibilistic Processes

Finally, Fig. 6.3 shows the class hierarchy of possibilistic processes. Poss_Dist has been specified above.

Transit_Matrix — A possibilistic transition matrix $\boldsymbol{\Pi}$.

| | | |
|---|---|---|
| **Data Attributes** | Universe | Integer $n = \|\Omega\|$ |
| | Data | A list of Poss_Dists $\left\langle \boldsymbol{\Pi}^{(j)} \right\rangle$. |
| **Invariant** | | $1 \leq j \leq n$ |

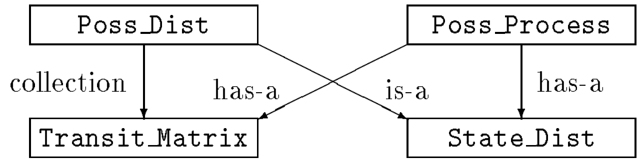


Figure 6.3: Possibilistic processes.

`State_Dist` — The current state possibility distribution $\pi^t$.

| **Data Attribute** | `Time` | The current time, an integer $t$ |
|---|---|---|
| **Invariant** | | $0 \leq t$ |

`Poss_Process` — A possibilistic process $\mathcal{Z}_\pi$.

| | | |
|---|---|---|
| **Method** | `Advance` | Determine next state function $\pi^t = \pi^{t-1} \circ \boldsymbol{\Pi}$. |

## 6.4 Extensions to the Basic Architecture

Of course, the architecture described above is merely the core for a broader implementation of possibilistic models, which must also involve a variety of measurement methods and links into CAST implementations of other aspects of GIT, let alone the input/output routines necessary for any software system.

### 6.4.1 Engineering Efficiencies and Algorithms

Some researchers have considered computationally efficient representations and algorithms for GIT.

#### 6.4.1.1 The Bayesian Approximation

Probabilistic approximations of non-specific random sets have been suggested. One is the maximum entropy probability distribution $p^{\mathcal{S}}$ (2.122) discussed in Sec. 2.6.4.2.

Voorbaak has also suggested the following specific approximation to a general random set.

**Definition 6.1 (Bayesian Approximation)** [293] Given a random set $\mathcal{S}$, let $\overline{m}\colon 2^\Omega \mapsto [0,1]$ be its **Bayesian approximation**, where $\forall A \subseteq \Omega$,

$$\overline{m}(A) := \begin{cases} \dfrac{\sum_{A_j \supseteq A} m_j}{\sum_{j=1}^N m_j |A_j|}, & |A| = 1, \\ 0, & |A| \neq 1 \end{cases} .$$

**Corollary 6.2** Given a random set $\mathcal{S}$, then $\overline{m}$ is an evidence function, and $\overline{\mathcal{S}}$ is specific with probability distribution $\overline{p}$ where $\forall \omega_i \in \Omega$,

$$\overline{p}(\omega_i) = \overline{p}_i := \overline{m}(\{\omega_i\}) = \frac{\text{Pl}_i}{\sum_{i=1}^{n} \text{Pl}_i}.$$

**Proof:** Since $\overline{m}(A) > 0$ only for non-singleton $A$, therefore we need only consider the values of $\overline{p}$ defined on the elements of $\Omega$. From the definition of evidence function (2.52), we need to show that

$$\sum_{i=1}^{n} \overline{p}_i = \sum_{i=1}^{n} \frac{\sum_{A_j \ni \omega_i} m_j}{\sum_{j=1}^{N} m_j |A_j|} = \frac{\sum_{i=1}^{n} \sum_{A_j \ni \omega_i} m_j}{\sum_{j=1}^{N} m_j |A_j|} = \frac{\sum_{j=1}^{N} m_j |A_j|}{\sum_{j=1}^{N} m_j |A_j|} = 1,$$

so that it is obvious that $\overline{p}$ is a probability distribution, $\overline{m}$ is an evidence function, and $\overline{\mathcal{S}}$ is a specific random set. The final result then follows from the plausibility assignment formula (2.68) and the lemma (2.69). ■

Voorbaak introduced this approximation for its computational efficiency in Dempster-combination problems, because it is invariant under the Dempster-combination operation $\odot$.

**Proposition 6.3** [293] Given two random sets $\mathcal{S}^1, \mathcal{S}^2$, then $\overline{m}^1 \odot \overline{m}^2 = \overline{m^1 \odot m^2}$.

However, this result is not crucial for this work, since Dempster combination is not crucial for possibility theory or process theory.

#### 6.4.1.2 The Möbius Transform

Of more importance is the so-called **Möbius transform** [143] or **fast Möbius transform** [284], which is an algorithm utilizing the Möbius inversion formula (2.59) to calculate among belief measures and evidence functions.

**Proposition 6.4 (Fast Möbius Transform)** [143, 284] Let the $\omega_i \in \Omega$ be taken in some arbitrary order, and assume a random set $\mathcal{S}$.

1. Assume the evidence function $m$ of $\mathcal{S}$, and let $m_0 := m$. Then $\forall A \subseteq \Omega$, and $1 \leq i \leq n$, determine $m_i$ by the algorithm

$$m_i(A) := \begin{cases} m_{i-1}(A) + m_{i-1}(A - \{\omega_i\}), & \omega_i \in A \\ m_{i-1}(A), & \omega_i \notin A \end{cases}.$$

   Then $m_n = \text{Bel}$, where Bel is the belief function of $\mathcal{S}$.

2. Assume the belief function Bel of $\mathcal{S}$, and let $m_n :=$ Bel. Then $\forall A \subseteq \Omega$, and $n \geq i \geq 1$, determine $m_i$ by the algorithm

$$m_{i-1}(A) := \begin{cases} m_i(A) - m_i(A - \{\omega_i\}), & \omega_i \in A \\ m_i(A), & \omega_i \notin A \end{cases}.$$

   Then $m_1 = m$, where $m$ is the evidence function of $\mathcal{S}$.

The fast Möbius transformation is extremely efficient, and will be used in implementing the relation

$$\texttt{Random\_Set} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Pl\_Measure}$$

from Fig. 6.1.

### 6.4.2 Supplementary Methods

A number of procedures have been described in the preceding chapters which are special or supplementary to the basic procedures, but which it would nevertheless be beneficial to implement explicitly. Most of these are transformations of one of the classes to another. The following relations can be appended to the basic diagrams above as appropriate.

**Distribution Operations:**

$$\texttt{Poss\_Dist} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Poss\_Measure}, \qquad \texttt{Prob\_Dist} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Prob\_Measure}.$$

From corollary (2.24), given a possibility distribution $\pi$, a possibility measure $\Pi$ can be constructed; similarly, from (2.83), given a probability distribution $p$, a probability measure Pr can be constructed.

**Probabilistic Approximations:**

$$\texttt{Random\_Set} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Prob\_Dist}.$$

Probability distribution approximations of random sets are available either as the maximum entropy probability distribution $p^{\mathcal{S}}$ (2.122), or the Bayesian approximation $\bar{p}$ (6.1).

**Compatibility Measures:** Given a `Prob_Dist` $p$ and `Poss_Dist` $\pi$, a variety of compatibility measures $\gamma(\pi, p)$ can be calculated (Sec. 2.6.3.3).

**Consonant Approximation:**

$$\texttt{Consistent\_RS} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Consonant\_RS}.$$

From (2.128), a consonant random set $\mathcal{S}^{\pi}$ can be constructed from a consistent random set $\mathcal{S}$.

**Frequency Conversions:**

$$\texttt{Prob\_Dist} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Poss\_Dist}, \qquad \texttt{Poss\_Dist} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Prob\_Dist}.$$

Probability and possibility distributions are co-determining from the methods presented in Sec. 3.4.2.2: maximum normalization ((3.16) and (3.17)), log-interval scaled UIP ((3.21) and (3.23)), probabilistic difference (3.25), and probabilistic nests (3.28).

**Possibilistic Normalization:**

$$\texttt{Random\_Set} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Consistent\_RS}, \qquad \texttt{Pl\_Assignment} \stackrel{\text{has-a}}{\longrightarrow} \texttt{Poss\_Dist}.$$

A consistent random set can be constructed from a random set $\mathcal{S}$, and a possibility distribution $\pi$ from a plausibility assignment $\vec{\text{Pl}}$, either by dimensional extension ((2.150) and (2.152) respectively), or by focused consistent transformation (assuming a given $i$, where $1 \leq i \leq n$, from (2.142) and (2.143) respectively). In the focused consistent transformation case, various methods from Sec. 2.8.1.2 to choose $i$ (maximum plausibility (2.144), minimal information distortion (2.148) and alternate minimal information distortion (2.149)) can be implemented and compared.

**Measurement:** The result of all the measurement methods presented in Chap. 4 is the construction of a random set, hopefully consistent. Therefore, while they require explicit implementation, they fall outside of the regular class hierarchy which has one root in the class `Random_Set`.

**Automata** Implementations of possibilistic processes can be extended to automata, as described in Sec. 5.6.1.

### 6.4.3 Other Extensions

There are further extensions which can be made to link the implementation of possibilistic methods with other GIT methods and other CAST implementations. These extensions can be considered either as a part of this research program, or as extensions to the research programs which have been or may be launched by others.

- Bernard de Baets of the University of Ghent is currently[2] implementing some of the methods described in Sec. 2.1 and Sec. 2.2.1 which generalize possibility theory beyond $[0, 1]$ to valuation on general lattices $\mathcal{L}$. There is therefore the opportunity to integrate these approaches.

[2]Personal communication.

- The origin of the CAST program is with the implementations of deterministic finite state machines of Pichler and his colleagues [213]. There is a clear relation to the possibilistic approach described here, and the opportunity to generalize to a variety of different GIT-based representations of finite state machines with uncertainty, including nondeterministic and stochastic machines.

- There has been some work [28, 81, 194] on the implementation of Klir's GSPS system [145]. It would certainly be very valuable to relate these efforts directly.

- In addition to the measurement methods described in Chap. 4, the possibilistic clustering methods from Sec. 3.4.4.3, including possibilistic $c$-means and the mountain method, can be integrated.

- While we have criticized the traditional dependence of possibility theory on fuzzy theory, the relationship is certainly worth maintaining. And although it is not our specific focus, there is value in relating possibilistic implementations with those of the variety of fuzzy set operations and concepts listed in Sec. 2.4.2, along with many others. There is by now a huge literature on these methods (see Kosko [165] and Terano, Asai, and Sugeno [283] for just two examples), and many academic and corporate efforts to develop fuzzy theoretical systems. Hopefully other researchers are building CAST-based implementations of fuzzy systems methods, which could then be integrated into this specifically possibilistic system. In particular, the action of conorm processes is just a case of fuzzy relation composition, so a generalized fuzzy CAST system implementing fuzzy relations can be used as a base for possibilistic automata.

- The architecture presented above (for example the class `Random_Set`) is defined on a discrete universe, while (as discussed in Sec. 4.1.6) the measurement methods of Chap. 4 are generally used on continuous universes. But also as discussed in Sec. 4.1.6, finite random sets on continuous universes can be redefined as random sets on discrete universes. Nevertheless, there are opportunities to reengineer this architecture in terms of continuous universes, and thereby make more general representations of fuzzy numbers, possibilistic histograms, and their continuous approximations, as discussed in Chap. 4.

- Finally, possibility measures, as extreme plausibility measures, exist within the more general Dempster-Shafer evidence theory. Therefore the extension of possibility theoretic implementations to include necessity measures (as the extreme belief measures dual to possibility measures), and general belief/plausibility pairs, may be very useful. Beyond that, of course, both belief and plausibility measures are special fuzzy measures, and so the ultimate extension is to the construction of CAST-based systems for fuzzy measures in general.

## 6.5 Empirical Investigations

As discussed at the beginning of this chapter, it is common in systems theory that computer-based implementation and simulation are necessary in order to investigate the properties of the systems under consideration, and this is the case with possibilistic systems, processes, and models. There are a number of issues which would be desirable to investigate empirically.

**Nonspecificity Calculations:** Determination of informational properties, and in particular nonspecificity values and the changes in these values, is of great interest. This would include, for example:

- Calculation of $\mathbf{N}(\pi^t)$ of the state vector of a possibilistic process as a function of $t$;
- Calculation of $\mathbf{N}(\pi)$ where $\pi$ is a possibilistic histogram, and the dependence of $\mathbf{N}(\pi)$ on the measurement method used;
- The change from $\mathbf{T}(\mathcal{S})$ to $\mathbf{T}(\pi)$ under normalization, and the dependence on both the general normalization method chosen from Sec. 2.8 and the various sub-choices required within some of the methods (for example, the choice of focus $\omega^*$ for a focused consistent transformation (Sec. 2.8.1.1)).
- Determination of $\mathbf{N}(\pi)$ under the cases where $\pi$ is a special fuzzy number, for example a parallelogram or triangle.

**Possibilistic Processes:** Aside from nonspecificity calculations, there are a number of other properties of possibilistic processes which require empirical investigation, for example the dependence of the form of possibilistic processes on the choice of norm operator $\sqcap$ and the choice of conditional possibility measure.

**Measurement Methods:** It is natural to explore the properties of the various measurement methods presented in Chap. 4 empirically, comparing the results of one data source using multiple methods.

**Uncertainty Invariance Transformations:** The UIP has already been mentioned in Sec. 6.4.2 as a frequency conversion method to be implemented as a supplementary method to the basic possibilistic architecture. While Klir and Parviz [158,160] have begun to empirically examine some results of frequency conversion methods, including the UIP, there are still many unanswered questions about the UIP. In the context of this work, it would be interesting to compare the time evolution of similar stochastic and possibilistic processes, and then compare those against their respective UIP transformations.

# Chapter 7

# Application to Model-Based Diagnosis and Trend Analysis of Spacecraft

Ground control to Major Tom: your circuit's
dead, there's something wrong.
— *David Bowie*

As possibility theory exists within the wider domain of GIT and Systems Science in general, so possibilistic models exist within a wider domain of modeling methodologies which utilize imprecision and uncertainty. These so-called **qualitative modeling** (QM) methods include traditional stochastic models, general fuzzy models, and others outside the range of GIT. Possibility theory promises to provide an important new approach within the QM movement.

Qualitative methods are appropriate for modeling complex systems, where the interaction among the large number of parts and varying environmental conditions results in the possibility of unpredictable behavior and long-run departure from established steady-state domains. Therefore one of the important applications of qualitative models is in the **model-based diagnostic** (MBD) approach to the identification of faults in complex systems such as spacecraft. These methods invoke multiple models of and measured data from the system in order to produce a set of candidate components for which there is a high confidence of failure.

This chapter examines the potential for the application of possibilistic models to the fault diagnosis and **trend analysis** of spacecraft systems.

## 7.1 Qualitative Modeling

**Qualitative Modeling** (QM), usually considered a part of artificial intelligence, can be broadly described as the attempt to deliberately model systems at a high level of abstraction from the actual systems themselves. Of course this approach produces models which are less precise than they might be, but with the tradeoff of potentially greater tractibility and accuracy (the less you say, the better your chance of being right).

The idea of reducing specificity in order to gain other important qualities in our models has long been a motivation for the use of GIT methods in Systems Science.

> As the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics. [322]

Thus QM as a whole is well within the tradition of GIT, and many of the specific methods of QM are directly related to GIT methods.

QM methods can be useful when there is only a poor model of the original system, or when there are missing or incomplete data. This can happen when systems are incompletely specified, when they have parameters or states which aren't always known with certainty, or when complexity makes detailed prediction difficult. For more information about QM in general, see the anthologies edited by Bobrow [16], and Fishwick and Luker [88], and survey articles by Fishwick [84] and Guariso, Rizzoli and Werthner [108].

There are a variety of broad approaches within QM, which come under the names of naive physics, qualitative physics, qualitative simulation, qualitative reasoning, qualitative dynamics, etc. There are also a number of specific methods, including bond graphs, causal loop modeling, natural language modeling, "lumped" state space models, and inductive approaches.

Here we are most interested in QM methods which use uncertainty distributions on state variables, and mixed interval- and point-valued dynamical systems. Models using uncertainty distributions are familiar from the discussion in Sec. 3.1.5, where both stochastic and possibilistic models were first introduced. In these methods the uncertainty about some attribute is represented mathematically by weights on all possible values. The set of weights, as a distribution, acts as a meta-state in the space of all possible distributions, and functional equations relating these meta-states produce predictions about the distribution meta-state at future times.

Models using probability distributions are also quite familiar (for example, Markov processes), and their correlates in possibility theory (for example, possibilistic automata) were introduced in Chap. 5. These methods are actually semi-qualitative, since the numerical representation of the distribution adds a quantitative component.

In an interval-valued dynamical modeling system like QSIM [171], a precise point-valued dynamical system of differential or difference equations is replaced by a homomorphic interval-valued process. Typically qualitative variables are identified within certain intervals, some relatively unconstrained (for example $x \in [0, \infty)$), and some constrained by "landmark values" (for example $x \in [x_{\min}, x_{\max}]$).

Qualitative variables are then generally related in three ways:

**Functional:** For example, if $y = M^-(x)$ then $y$ is a monotonically decreasing function of $x$, so that if $x \in [0, x_{\max})$ then $y \in (-\infty, 0]$ or $y \in (M^-(x_{\max}), 0]$.

**Arithmetic:** Standard mathematical operations can also be represented qualitatively, for example if $x \in [0, x_{\max}]$ and $y \in (-\infty, 0]$, then $xy \in (-\infty, 0]$, but $x + y$ is unknown.

**Dynamic:** Change of state is represented by qualitative magnitude and direction. Qualitative differential relations link directions with magnitudes, for example given $y = dx/dt$, then

$$x \text{ increasing} \rightarrow y \in (0, \infty), \qquad x \text{ decreasing} \rightarrow y \in (-\infty, 0).$$

Of course, determinative results may not be available in such a qualitative model. For example, we saw above that $x + y$ could be any value in $(-\infty, \infty)$. Similarly, the existence of landmark values leads to uncertainty as to whether a landmark has been crossed. To account for each possibility, two alternatives must be branched off. Therefore in general, QM systems have a tree of possible system behaviors, and external factors (heuristics or other constraints) may be required to prune that tree.

## 7.2 Model-Based Diagnosis (MBD)

A serious problem for NASA-Goddard is the diagnosis of faults in complex systems (spacecraft) given only the knowledge of their outputs (telemetry data). **Trend analysis** is similar to long-range diagnosis, in which, given the persistence of a potentially anomalous condition, and assuming continuation of the trend, the goal is to predict future states and failure modes.

There are many approaches to diagnosis and trend analysis. The model-based approach (MBD) [116] is based on the premise that knowledge about the internal structure of a system can be useful in diagnosing its failure. In MBD, a software model of the system, given inputs from the real system, generates and tests various failure hypotheses.

A typical MBD approach (derived from some of the standard literature [45, 77, 113]) to diagnosing a spacecraft (here described as some internal **system** whose **sensor** measurements output to a **telemetry** stream) is shown in Fig. 7.1.

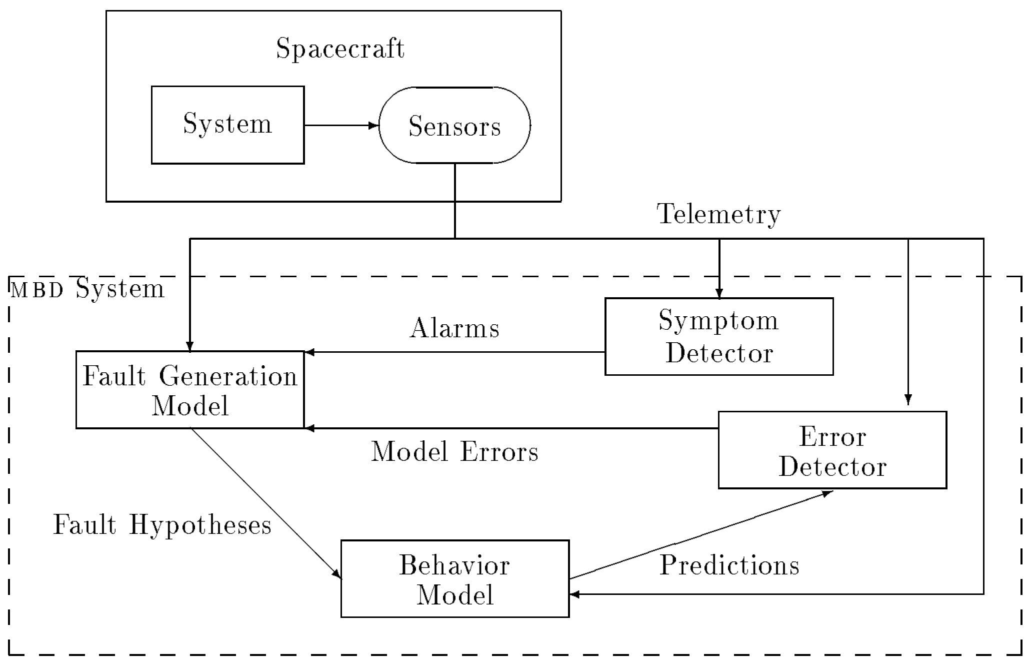


Figure 7.1: A typical model-based diagnostic system.

An **alarm** is a report that some observed system attributes have departed from nominal, and entered error, conditions, usually by exceeding some threshold values; a **prediction** is a report that some system attributes should be in certain states; a **fault hypothesis** is a list of system components which may have failed; and an **error** is a report of a discrepancy between predicted and measured system attribute states.

The overall MBD system then involves two distinct spacecraft models. The **fault generation model** (FGM) takes inputs from telemetry, alarms, and errors, and either produces anew, or modifies existing, fault hypotheses. The **behavior model** takes inputs from telemetry and fault hypotheses, and outputs predictions. These are then corroborated against telemetry to produce errors. The fault hypotheses act to modify the behavior model so that it predicts system behavior as if the hypothetical system components had actually failed.

Both models can be difficult to construct, typically involving delicate tradeoffs among accuracy, precision, and tractibility. But the FGM, as the heart of the MBD approach, is particularly complex and involved. The FGM could be, for example, an inversion of the behavior model (as for Dvorak and Kuipers [77]) or a decision tree (as for Shen and Leitch [265]). Through backwards reasoning a variety of subsets of components can be identified, any of which are consistent with the given telemetry and alarms.

**Filtering** is the process by which error output is used to prune the set of fault hypotheses. If the prediction of the behavior model as modified by a particular fault hypothesis produces errors, then that fault hypothesis is not retained. As the system is monitored over time, further observations narrow the class of fault hypotheses. Achieving the null set indicates model insufficiency. But if the overall MBD system stabilizes to a non-empty set of fault-hypotheses, then these are advanced as possible causes of the failure.

QM has been applied to MBD to produce qualitative model-based diagnostic systems. For example, in the approach of Dvorak and Kuipers [77], model predictions are intervals of possible system state values. Stochastic methods, for example Bayesian networks [96] and Markov processes [107], have been used extensively in MBD applications. And recently Shen and Leitch [265,266] have advanced the FuSim method for qualitative MBD which uses fuzzy arithmetic.

## 7.3 Possibility Theory as a Qualitative Modeling Method

As discussed in Sec. 3.3.6, possibilistic models my be appropriate where stochastic concepts and methods are not, including situations where long-run frequencies are difficult if not impossible to obtain, or where small sample sizes prevail. This is true in **reliability analysis**, for example, where failures and system entry into non-nominal behavior domains are very rare; and trend-analysis, where even though observations are made over a long time, the state variables of concern change only very slowly, and new domains of behavior are only very rarely seen. In these cases the weakness of the possibilistic representation is matched by the weak evidence available.

There are many reasons why it can be hoped, and even expected, that possibility theory can come to play an important role in QM in general, and in the application of QM to MBD in particular.

Hamscher et al. have noticed some of the weaknesses of stochastic methods for MBD.

> It is usually assumed that reliable failure statistics will be available, but this is in fact rare in practice. What is needed ... is a way of working with likelihoods that could be specified ordinally rather than quantitatively. [116, p. 452]

This is *exactly* what possibility theory provides, a non-additive, ordinal approach to QM which hybridizes interval-valued dynamics and uncertainty distribution methods.

From (2.46), and under the definition

$$\pi^\alpha := \begin{cases} \{x \in \mathbb{R} : \pi(x) \geq \alpha\}, & \alpha \in (0,1] \\ \mathbf{U}(\pi), & \alpha = 0 \end{cases},$$

$\pi$ can be represented as a set of nested intervals $\pi^\alpha$ weighted by the possibility values $\alpha$ (see Fig. 7.2). Thus as discussed in Sec. 2.4.3, fuzzy intervals, for example possibilistic histograms and their continuous approximations, and the mathematical operations on them, generalize the methods of interval analysis [187].

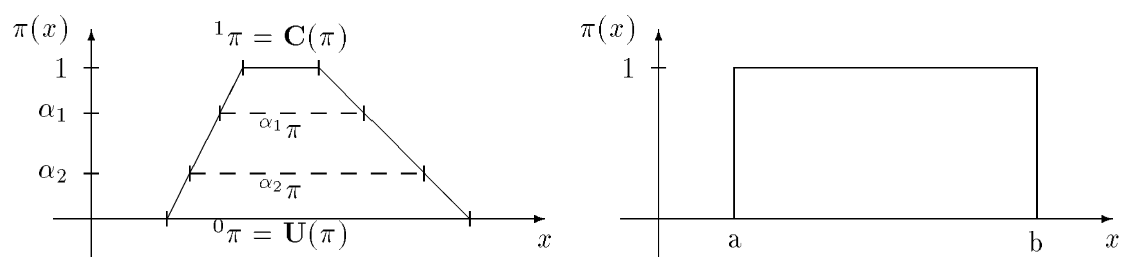


Figure 7.2: (Left) A possibility distribution as a collection of weighted intervals. (Right) The special case of a crisp interval.

Fuzzy arithmetic methods are very popular, and because all fuzzy intervals and numbers as in fact possibility distributions, therefore QM methods which use fuzzy arithmetic are essentially possibilistic (for example the recent work of Sugeno and Takahiro [278]). Fuzzy arithmetic has been used as a QM method for MBD, for example by Shen and Leitch [266] and Fishwick [87]. They use the standard methods of fuzzy control systems, where a set of overlapping fuzzy intervals divide a quantity space into a few linguistic values like "large positive" and "small negative". These fuzzy sets are not measured properties of the system being modeled, and are dependent on the heuristic specification of the system modeler. Thus, as critiqued in Sec. 3.4.1.2, they are essentially modeling the cognitive state of some human expert, rather than directly modeling the system in question.

This contrasts sharply with QM methods based on possibilistic processes. First, they are cast strictly within the context of mathematical possibility theory (including possibilistic processes) specifically, rather than fuzzy theory generally. Also, they are based on measurement of the system in question.

## 7.4 A Possibilistic Approach to MBD

At both the general level and in some specific ways, there are areas of MBD for which it is appropriate to consider a possibilistic approach.

### 7.4.1 Possibilistic Symptom and Error Detection

Typically the symptom and error detectors simply compare the measured value against a crisp interval of nominal or predicted values (for example in Dvorak and Kuipers [77]). This is inadequate because the resulting cutoff from nominal to error condition is essentially arbitrary. It is natural to use a fuzzy interval to generalize this, measuring either prediction errors or fault symptoms as the possibilistic distance of the telemetry from the predicted or nominal system state respectively.

Consider a measured value $x$ compared against an error fuzzy interval of the form of Fig. 7.2. Such a possibility distribution could be the output of the behavior model, for instance, and would then serve as input to the error detector. Then $\pi(x)$ is the strength of the error or alarm raised. When $x \in \mathbf{C}(\pi)$, then $\pi(x) = 1$ and there is no alarm. When $x \notin \mathbf{U}(\pi)$, then $\pi(x) = 0$ and the alarm is complete. In between, an intermediate alarm is raised.

Even in situations where crisp thresholds are acceptable, they may be dynamic, varying as a result of changing system and environmental conditions. Doyle et al. consider the situation of an earth-orbiting spacecraft as it proceeds through sunlight and shadow.

> Impingent solar radiation changes the thermal profile of the spacecraft, as does the configuration of currently active and consequently, heat-generating subsystems on board. Thresholds on temperature sensors should be adjusted accordingly. A particular temperature value may be indicative of a problem when the spacecraft is in shadow or mostly inactive, but may be within acceptable limits when the spacecraft is in sunlight or many on-board systems are operating. [51]

This situation is shown in Fig. 7.3. Assume a variable, say the temperature $t$ of a given component, must be kept in a critical range as the spacecraft moves in and out of daylight. As it does so, the range shifts as shown in the upper figure, where the transition periods begin at a change in sunlight, and continue to thermal equilibrium. For simplicity, assume that that interval is sampled uniformly six times during the orbital day, twice each for daylight $D_i$, night $N_i$, and transition period $T_i$. The possibilistic histogram for the possibility $\pi(t)$ of $t$ holding a value at any given time and a parallelogram approximation are shown.

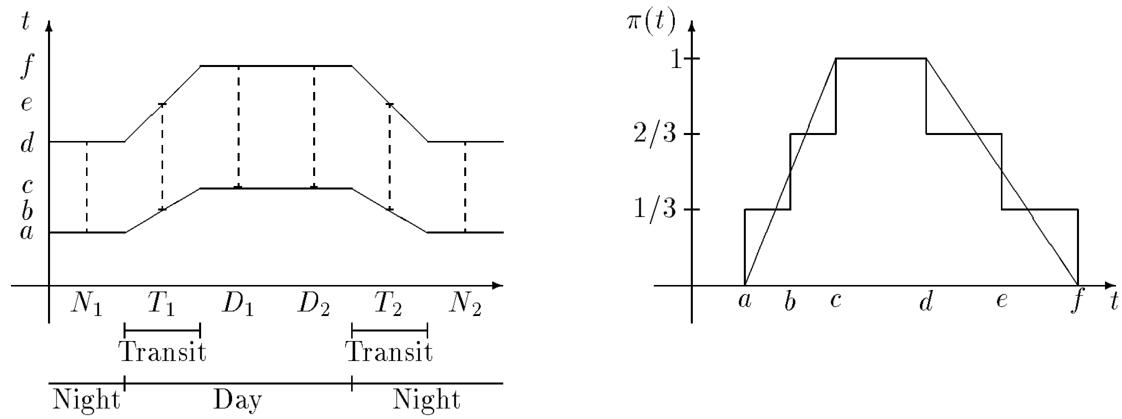


Figure 7.3: (Left) Variable critical range of a component through a day-night cycle. (Right) Its possibilistic histogram and a parallelogram approximation.

A combination of these two approaches is also possible, where instead of a crisp interval changing over time, rather a whole possibility distribution itself changes with time.

### 7.4.2 Sensor Modeling

Although the MBD system contains two models, the behavior model and the FGM, as a whole, it is itself also a model of the spacecraft. As such, it is dependent on its inputs from measurement, and thus on the sensor output of the spacecraft. Thus there are modeling issues in MBD concerning the sensors themselves.

When modeling complex systems, sensor data may be sparsely distributed, with missing observations, and sometimes very small samples sizes. As discussed in Sec. 3.3.6, these are important conditions for the inapplicability of stochastic methods, and when they hold, possibilistic methods should be considered.

In this respect, there is strong support in the literature for the idea that possibility, as distinct from probability, has a role to play in QM. For example, Luo and Kay observe

> When additional information from a sensor becomes available and the number of unknown propositions is large relative to the number of known

> propositions, an intuitively unsatisfying result of the Bayesian approach is that the probabilities of known propositions become unstable. [180]

While Durrant-Whyte takes a typical statistical approach, he also notes

> A robot system uses notably diverse sensors, which often supply only sparse observations that cannot be modeled accurately. [76]

Dvorak and Kuipers make a similar observation in the context of model-based monitoring.

> All measurements come from sensors, which can be expensive and/or unreliable and/or invasive. Monitoring is typically based on a small subset of the system parameters, with limited opportunity to probe other parameters. [77]

#### 7.4.2.1 Data Fusion

Possibilistic measurement, as outlined in Chap. 4, is predicated on the observation of subsets or intervals which are partially overlapping. It is therefore imperative to consider the source of these intervals, some of which are described in Sec. 4.5.1.4. All of these conditions arise in MBD when considering the problem of **data fusion** [180], that is the combination of data from *different* instruments which measure the same system attribute, either directly or indirectly.

Hackett and Shah discuss data fusion in general, including indirect measurements, and the Dempster-Shafer (that is, random set) approach.

> Every sensor is sensitive to a different property of the environment; in order to sense multiple properties, it is necessary to use multiple sensors. A system using multiple sensors that sense a single property can be used. [110]

Dubois, Lang, and Prade [53] have also considered the data fusion problem using possibilistic logic.

**Indirect Measurements** First consider the situation where measurements of a component are not made directly, but rather knowledge of the state of the component is only gained indirectly by inference from the outputs of sensors of other components. Doyle et al. [51] offer an example from jet aircraft: low engine thrust can be indicated by either low exhaust temperature or low turbine rotation speed, or both.

This situation is illustrated in Fig. 7.4. Here component A is not monitored. Its state can only be inferred from the sensors D and E, which monitor components B and C, and which in turn are causally connected to A. Each of the intervals reported by D and E individually is distinct and disjoint. But since the knowledge of A provided by D and E is mediated by B and C, together they may indicate that A exists in two different, possibly overlapping, intervals.

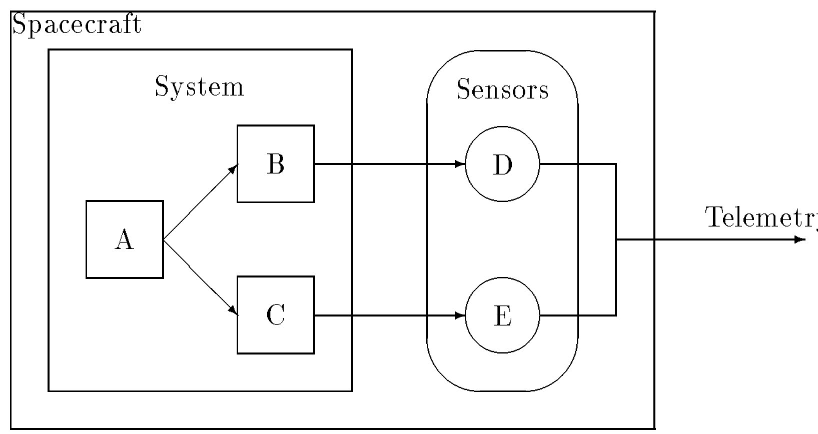


Figure 7.4: Indirect measurements of the state of a spacecraft component.

So as the amount of sensor "penetration" (sensor/component ratio) drops, standard measurement methods yielding frequency distributions may become less tenable, leaving only observations of random sets.

**Redundant Measurements** Alternatively, a system component may be monitored redundantly by multiple instruments. If these sensors are identical, and identically calibrated, then the result will simply be as if there was a time-series of observations from a single instrument. But if they are mutually discalibrated, either out of phase, or scale, or both, then the intervals reported from each instrument may overlap.

If the sensors measure distinct modalities (e.g. pressure and temperature) of a single component, then a process of **registration** [110] is required to derive a report from one in the modality of the other, or two new reports from each in a third modality. In any event, the argument here is very similar to the one above in the case of indirect measurements, and possibly overlapping intervals my result.

#### 7.4.2.2 Sensor Failure Modeling

As mentioned above, in MBD data are not only combined from disparate sensors, they are also sometimes incomplete, degraded, or missing altogether. Even when standard (disjoint) observations are made, under these conditions there is the potential for the application of GIT and possibility theory.

First, as discussed in Sec. 4.1.5, in GIT missing observation are represented as observations of the *entire* universe $\Omega$, resulting in empirical random sets from augmented specific measuring devices $\mathcal{C}^+$. While this does not result in a specifically possibilistic situation, neither does it result in a frequency or probability distribution.

When sensor data are not missing, but rather degraded, compromised, or suspect in some way, a confidence weighting on each sensor's output is naturally not additive: our confidence about the sensors is not divided among them, since all could be perfect or any number of them could be in any state of degradation. Instead it is natural to represent this confidence as a possibility distribution on each sensor's output. Again, an observation in the core indicates complete confidence, while one outside the support indicates complete sensor failure.

Representation of a graduated degree of sensor failure allows a corresponding graduated degree of confidence in model predictions. The need for this has been noted by Fulton.

> When we detect a broken sensor, great difficult arises if we continue diagnosing other failures, because typical rule-based systems do not degrade gently when sensors fail (because the mapping is dependent on a complete and accurate set of sensor data). [89]

### 7.4.3 Possibilistic Models Proper

In the sequel, the term **system model** will refer to the FGM or the behavior model generally. So finally, it is useful to consider possibilistic methods applied directly to the system models themselves, constructing them as possibilistic processes such as possibilistic automata, and not as fuzzy arithmetic systems as discussed in Sec. 7.3.

Input to these systems may or may not be proper possibility distributions, since both crisp (standard) intervals and point values are special cases of possibility distributions. But if they are, then it was discussed how telemetry, alarms, and errors can be possibilistically weighted. A possibilistic FGM then would be responsible for producing as its output a set of fault hypotheses which are possibilistically weighted for input to the behavior model. This would in turn generate model prediction errors with possibilistic weights.

A system model which is a possibilistic automata can also be cast as a possibilistic Markov process. The semantics of the transition matrix $\boldsymbol{\Pi}$ in a system model is understood in terms of a subsystem-level model where the conditional possibilistic weight indicates a non-additive coupling or relatedness among subsystems. This could be, for example, the efficiency of the subsystem, as in the approach of Doyle et al. [51]. Or, when considering the system model as a causal graph, as in the approach of Hall et al. [113], the weights indicate the degree of causal connectivity between subsystems.

Thus in the possibilistic approach a system model is essentially a possibilistic network, where nonadditive, possibilistic weights are placed on the arcs of a causal graph. The corresponding network appears similar to a Bayesian network, but the mathematics is possibilistic, not stochastic.

# Appendix A

# Original Contributions

This dissertation makes the following original contributions in each of the following chapters:

1. **Introduction: Science and Information:** It is argued that while classical information theory has historically been closely linked to the physical sciences, the new Generalized Information Theory (GIT) is decidely not so linked.

2. **Mathematical Possibility Theory:** Both a review of GIT and some original developments are provided.

    - Sec. 2.2.1, p. 19: A new **algebraic axiomatization** of possibility is provided.
    - Sec. 2.3, p. 22: **Distributions of fuzzy measures** are introduced to provide a link between fuzzy measures and fuzzy sets through random sets; and the **historical relation** between the terms "fuzzy set" and "fuzzy measure" is explicated.
    - Sec. 2.5.2, p. 32: **Distributions on random sets**, as well as their **operators**, their **completion**, and their **structural** and **numerical aggregation** functions are introduced; the special cases of **probability** and **possibility** are derived.
    - Sec. 2.7.3.1, p. 49: It is proved that the **optimal approximation** to a consistent random set is the **unique consonant approximation**, and that
    - Sec. 2.7.3.4, p. 51: it is also the **canonical random set** which is one-point equivalent to a given maximally normalized fuzzy set.

- Sec. 2.8, p. 51: Methods for **possibilistic normalization** of possibility distributions, and thus for constructing **possibilistic approximations** of random sets, are defined and developed, including **dimensional extension** and **focused consistent transformation**.
- Sec. 2.8.1.2, p. 55: **Focus selection methods** for focused consistent transformations are described, including **maximum plausibility** and **minimal information distortion**.
- Sec. 2.8.3, p. 61: **Other methods** for consistent approximations are rejected, including **maximum compatibility** with the one-point probabilistic approximation and **maximum entropy** of the evidence function.
- Sec. 2.9, p. 61: It is argued that possibility theory is **distinct and autonomous** from fuzzy theory, and that there is **no special relationship** between fuzzy theory and possibility theory. The argument is made on the basis of **possibilistic normalization**, the **ubiquity of fuzzy sets** in GIT, the relation between **alpha cuts and focal elements** of random sets, and the relation between the **possibilistic operator** and **fuzzy unions**.
- Sec. 2.10, p. 72: It is argued that possibility theory is **formally independent** of but **weakly related** to probability theory.

3. **Possibilistic Semantics:** Development of the fundamental conceptual categories of possibility theory.

    - Sec. 3.1, p. 76: It is argued that a **semantics** of **graduated,** ***de re*** possibility must be derived from the perspective of **semiotics** and **general modeling relations**, where interpretations are constrained, but not determined, from possibilistic mathematics alone.
    - Sec. 3.2, p. 82: It is argued that possibilistic semantics must be embedded in the contexts of the **modal, logical, stochastic**, and **natural language** concepts of possibility.
    - Sec. 3.2.4, p. 86: It is argued that the **observation** of an event must be identified with its **complete possibility**.
    - Sec. 3.2.5, p. 87: A **strong compatibility** criterion for probability and possibility is advanced and developed.
    - Sec. 3.3, p. 94: The **conceptual basis** for the interpretation of possibility statements is developed with respect to **mathematical possibility theory**, **possibilistic processes**, **complex systems**, and **physical systems**.

- Sec. 3.4, p. 108: **Existing semantics** of possibility are **critiqued**, including:
  - Sec. 3.4.1, p. 108: The insufficiency (and predominance) of **subjective evaluation** methods;
  - Sec. 3.4.2, p. 118: **Conversion of frequency or probability distributions** to determine possibility distributions, including **compatibility calculations** of the various conversion methods;
  - Sec. 3.4.3, p. 124: Interepretation of possibilities as **statistical likelihoods**;
  - Sec. 3.4.4, p. 128: Existing **objective measurement** methods for possibility theory.

4. **Possibilistic Measurement:** Definition and development of possibilistic measurement methods.

- Sec. 4.1, p. 132: It is argued that **set statistics** (interval data) are the only legitimate basis for possibilistic measurement.
- Sec. 4.1.3, p. 133: **Empirical random sets** are defined from **set-frequencies**.
- Sec. 4.1.5, p. 135: **Incomplete observations** are interpreted in the context of set statistics.
- Sec. 4.2, p. 137: **Possibilistic histograms** are defined and characterized as piecewise constant possibility distributions resulting from consistent empirical random sets.
- Sec. 4.2.3, p. 142: Possibilistic histograms are shown to be **fuzzy intervlas**.
- Sec. 4.3, p. 143: **Continuous approximations** of possibilistic histograms are developed.
- Sec. 4.4, p. 147: The **compatibility** of possibilistic histograms with probability distributions is explored.
- Sec. 4.5, p. 151: **Sources of set statistics** are considered, including:
  - Sec. 4.5.1, p. 151: **Ensembles** of multiple, differently calibrated instruments;
  - Sec. 4.5.3, p. 155: **Order statitics** on specific (point) data, including **focused intervals**, **interval cores**, and **consonant intervals**;
  - Sec. 4.5.4, p. 164: **Local extrema** of a point data stream.

5. **Possibilistic Processes:** Definition and development of possibilistic processes.

    - Sec. 5.1.1, p. 167: Introduction of **conorm semirings**, and the clarification of the relationship between **norm/conorm** operator pairs and semirings.
    - Sec. 5.1.2, p. 169: Use of conorm semirings in **fuzzy matrix composition**.
    - Sec. 5.2, p. 170: Introduction of **conorm processes** using conorm semirings.
    - Sec. 5.3.2.2, p. 176: Casting of **stochastic processes** in terms of conorm processes.
    - Sec. 5.3.2.4, p. 179: Introduction of **possibilistic processes** as **possibilistically normal** general fuzzy processes.
    - Sec. 5.4, p. 181: Derivation of **properties** of possibilistic processes.
    - Sec. 5.5.2.3, p. 191: The **strong compatibility** of possibilistic with stochastic processes is demonstrated.
    - Sec. 5.5.3, p. 192: **Conditional** fuzzy measures and distributions are defined in terms of conorm semirings, and in particular **conditional possibility** is defined, and some of its properties established.
    - Sec. 5.6.1, p. 197: **Possibilistic automata** are defined by extending possibilistic processes to include input and output functions.
    - Sec. 5.6.2, p. 198: A **possibilistic Monte Carlo methods** is defined and developed.
    - Sec. 5.6.3, p. 200: **Possibilistic Markov processes** are defined.

6. **Software Implementation:** Proposed architecture of C++ classes to implement possibilistic processes and models.

7. **Qualitative Model-Based Diagnosis:** Application of possibilistic measurement and processes to the qualitative modeling of a complex system.

    - Consideration of possibility theory in the context of other **qualitative modeling** methods.
    - Descripton of a possibilistic approach to **model-based diagnosis**, including:

- The use of **fuzzy intervals** for graduated **symptom** and **error detection**;
- The representation of **dynamic crisp error thresholds** by possibilistic histograms;
- The use of random set measurement in **data fusion** problems, including **indirect** and **redundant** situations;
- Possibilistic representations of **sensor degradation** and **sensor failure**;
- And the representation of **system models** as **possibilistic networks**.

# Appendix B

# Further Work

This dissertation leaves many open avenues of research, including both specific questions and general directions. Some of these are pointed out here.

**Possibility Theory and GIT:**

- Further development of lattice-valued possibility theory.
- The search for fuzzy measures other than probability and possibility which have distributions and structural and numerical aggregation functions.
- Proof of (2.129), p. 49.
- Development of other well-justified possibilistic normalization methods.
- Further exploration of minimal distortion possibilistic normalization, including the change in strife vs. nonspecificity levels and the affect of the choice of distortion function.
- Further exploration of the general nature of distributions of fuzzy measures in the context of random sets.

**Possibilistic Semantics:**

- Further research of the historical use of graduated and/or *de re* possibility in philosophy.
- Further consideration of the relation between modal logic and possibility theory.
- Further development of possibilistic concepts, especially with relation to the action and operation of real complex systems which can be seen as having elastic constraints.

- Further research on the interaction between fuzzy logic and quantum theory.
- Continued study of new frequency conversion methods with respect to compatibility.
- Exploration in particular of uncertainty invariance frequency conversion.
- Continued research on new possibilistic clustering methods as objective measurement techniques.

**Possibilistic Measurement:**

- Investigation of real measurement systems which produce subset or interval results.
- Further investigation of the empirical properties of possibility distributions determined from the different measurement methods presented in Sec. 4.5, for example order statistical data and local extrema.
- Comparison of the empirical properties, including the nonspecificity measures, of possibilistic histograms produced from the various continuous approximation methods presented in Sec. 4.3.

**Possibilistic Systems:**

- Empirical investigation of the behavior of possibilistic processes and automata, especially the development of their information measures over model time.
- Since measured possibility distributions are fuzzy numbers, it would be interesting to consider fuzzy arithmetic as a possibilistic process, in particular any general formulae for changes in nonspecificity with fuzzy arithmetic operations.
- Consideration of the existence of other conorm semirings.
- Establishment of the necessary, and not just sufficient, conditions for stochastic and possibilistic automata.
- Consideration of processes converted from stochastic to possibilistic forms through uncertainty invariance.
- Consideration of the properties of random set Monte Carlo selection specifically in the possibilistic (consonant random set) case.
- What are the conditions under which the fuzzy interval status of state vectors of possibilistic processes is preserved?

**Implementations:** The entire research program for the implementation of possibilistic systems and models described in Chap. 6 requires fulfillment.

**Model-Based Diagnosis:** The entire research program described in Chap. 7 also requires fufillment.

# Appendix C

# Related Publications

The following papers were presented and published in conjunction with and during the writing of this dissertation.

- "Hierarchy, Strict Hierarchy, and Generalized Information Theory," *Proceedings of the 1991 Conference of the International Society for the Systems Sciences,* Östersund, Sweden, v. 1, pp. 123-132, 1991. Winner, Vickers Memorial Award for Best Student Paper.

  Strict hierarchy as tree, loose hierarchy as DAG; loose structural hierarchies as class relations; proposed measure of looseness of random sets.

- "Towards an Empirical Semantics of Possibility Through Maximum Uncertainty," *Proceedings of the 4th World Congress of the International Fuzzy Systems Association: Artificial Intelligence*, Free University of Brussels, Belgium, pp. 86-89, 1991.

  Chap. 2: Extended abstract: application of maximum nonspecificity as a normalization method for possibilistic random sets.

- "Possibilistic Measurement and Set Statistics", in: *Proceedings of the 1992 Conference of the North American Fuzzy Information Processing Society*, v. **2**, pp. 458-467, 1992.

  Chap. 4: Set-based statistics to generate possibility distributions from measured data; physical measurements methods to generate statistical data.

- "Minimal Information Loss Possibilistic Approximations of Random Sets", with George Klir. In: *Proceedings of the 1992 IEEE Int. Conf. on Fuzzy Systems*, San Diego, IEEE, pp. 1081-1088, 1992.

Chap. 2: An empirical measuring procedure which yields data governed by possibility theory. Set-based statistics are used to generate empirically derived random sets. Normal possibility distributions are available for all consistent random sets, and a set of "consistent transformations" are available for all inconsistent random sets. The Principle of Uncertainty Invariance is used in a modified form to select the consistent transformation with minimal information loss from the original random set.

- "Possibilistic Semantics and Measurement Methods in Complex Systems", in: *Proceedings of the Second International Symposium on Uncertainty Modeling and Analysis*, University of Maryland, ed. Bilal Ayyub, pp. 208-215, IEEE Computer Society, 1993.

  Chap. 3: Towards development of a strictly possibilistic semantics of natural systems: the semantics of possibility statements in relation to modal logic, natural language, and mathematical possibility; strong consistency relation for probability and possibility; and the application of possibility theory to complex systems.

- "On Possibilistic Automata", in: *Computer Aided Systems Theory—EUROCAST '93*, ed. F. Pichler and R. Moreno-Díaz, pp. 231-242, Springer-Verlag, Berlin. Selected for publication in the select proceedings of the 1993 Computer-Aided Systems Theory (CAST) Conference.

  Chap. 5: Possibilistic automata as pessimistic fuzzy automata which are normal in the sense of general automata; properties of possibilistic automata; possibilistic automata are identical with strongly consistent stochastic automata.

- "Some New Results on Possibilistic Measurement", *Proceedings of the 1993 Conference of the North American Fuzzy Information Processing Society*, Allentown Pennsylvania, pp. 227-231, 1993.

  Chap. 4: Possibilistic histograms, their interpretation as fuzzy numbers, and their continuous approximations.

- "Empirical Possibility and Minimal Information Distortion", in: *Fuzzy Logic: State of the Art*, edited by R. Lowen and M. Roubens, Kluwer, pp. 143-152, 1993. Invited paper.

  Chap. 2: Normal possibility distributions are available for consistent random sets, and a set of focused consistent transformations is available for inconsistent random sets. The Principle of Uncertainty Invariance is modified to provide a method which selects that consistent transformation with Minimal Information Distortion from the measured random set.

- "Qualitative Model-Based Diagnosis Using Possibility Theory", to be presented at the 1994 Goddard Conference on Space Applications of Artificial Intelligence.

  Chap. 7: Possibility theory as a qualitative modeling method for the model-based diagnosis of spacecraft system faults: possibilistic symptom and model prediction error detection; sensor failure modeling and data fusion through indirect and redundant measurements; and fault generation and behavior models as possibilistic networks and nonadditive causal networks.

- "An Object-Oriented Architecture for Possibility Theoretic Implementations", to be presented at the 1994 Computer-Aided Systems Theory Conference.

  Chap. 6: Design for C++ classes for random sets, probability and possibility distributions, and possibilistic processes.

# Appendix D

# Summary of Possibility Theory

Table D.1 summarizes the main formulae of possibility theory.

| | RANDOM SET | DISTRIBUTIONS: $i \leftrightarrow j$ | |
|---|---|---|---|
| | | Probability | Possibility |
| Focal Set | Any | Singletons: $A_i = \{\omega_i\}$ | Nest: $A_i = \{\omega_1, \ldots, \omega_i\}$ |
| | | $\{\omega_i\} = A_i$ | $\{\omega_i\} = A_i - A_{i-1}, A_0 := \emptyset$ |
| Structure | None | Partition | Total order |
| Belief | $\mathrm{Bel}(A) = \sum_{A_j \subseteq A} m_j$ | $\Pr(A) := \mathrm{Bel}(A)$ | $\eta(A) := \mathrm{Bel}(A)$ |
| Plausibility | $\mathrm{Pl}(A) = \sum_{A_j \cap A \neq \emptyset} m_j$ | $\Pr(A) := \mathrm{Pl}(A)$ | $\Pi(A) := \mathrm{Pl}(A)$ |
| Relation | $\mathrm{Bel}(A) = 1 - \mathrm{Pl}(\overline{A})$ | $\mathrm{Bel}(A) = \mathrm{Pl}(A) = \Pr(A)$ | $\eta(A) = 1 - \Pi(\overline{A})$ |
| Distribution | $\mathrm{Pl}_i = \sum_{A_j \ni \omega_i} m_j$ | $p_i := \mathrm{Pl}_i = m_i$ | $\pi_i := \mathrm{Pl}_i = \sum_{j=i}^{n} m_j$ |
| | | $m_i = p_i$ | $m_i = \pi_i - \pi_{i+1}, \pi_{n+1} := 0$ |
| Measure | | $\Pr(A \cup B) = \Pr(A) +$ $\Pr(B) - \Pr(A \cap B)$ | $\Pi(A \cup B)$ $= \Pi(A) \vee \Pi(B)$ |
| Normalization | | $\sum_i p_i = 1$ | $\bigvee_i \pi_i = 1$ |
| Operator | | $\Pr(A) = \sum_{\omega_i \in A} p_i$ | $\Pi(A) = \bigvee_{\omega_i \in A} \pi_i$ |
| Nonspecificity | $\sum_j m_j \log_2 \lvert A_j \rvert$ | 0 | $\sum_{i=2}^{n} \pi_i \log_2 \left[\frac{i}{i-1}\right]$ |
| | | | $= \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i)$ |
| Strife | $-\sum_j m_j \log_2 \left[\sum_{k=1}^{n} m_k \frac{\lvert A_j \cap A_k \rvert}{\lvert A_j \rvert}\right]$ | $-\sum_i p_i \log_2(p_i)$ | $\sum_{i=2}^{n} \pi_i - \pi_{i+1} \log_2 \left[\frac{i^2}{\sum_{j=1}^{i} \pi_j}\right]$ |
| | | | $< .892$ |
| Semiring | $\langle \oplus, \otimes \rangle$ | $\langle +, \times \rangle$ | $\langle \vee, \sqcap \rangle$ |
| Marginals | | $p(x) = \sum_y p(x,y)$ | $\pi(x) = \bigvee_y \pi(x,y)$ |
| Conditionals | | $p(x,y) = p(x \vert y) \times p(y)$ | $\pi(x,y) = \pi(x \vert_{\sqcap} y) \sqcap \pi(y)$ |
| | | $p(x \vert y) = p(x,y)/p(y)$ | $\pi(x \vert_{\sqcap} y) \in [\pi(x,y), 1]$ |
| | | $\forall y, \sum_x p(x \vert y) = 1$ | $\forall y, \bigvee_x \pi(x \vert_{\sqcap} y) = 1$ |
| | | $\mathbf{P} := [p(x \vert y)]$ | $\boldsymbol{\Pi} := [\pi(x \vert_{\sqcap} y)]$ |
| Process | | $p' = p \cdot \mathbf{P}$ | $\pi' = \pi \circ \boldsymbol{\Pi}$ |
| | | $p'(x) = \sum_y p(y) \times p(x \vert y)$ | $\pi'(x) = \bigvee_y \pi(y) \sqcap \pi(x \vert_{\sqcap} y)$ |
| Concepts | | Division among distinct hypotheses | Coherence around certain hypotheses |
| | | Frequency | Capacity |
| | | Chance | Ease of attainment |
| | | Likelihood | Distance, similarity |

Table D.1: Summary of probability and possibility in GIT.

# Appendix E

# Mathematical Notation

| Symbol | Meaning |
|---|---|
| $\mathbf{0}$ | Boolean algebra infimum |
| $\mathbf{1}$ | Boolean algebra supremum |
| $\vec{1}_i$ | Certain distribution |
| $2^\Omega$ | Power set of $\Omega$ |
| $[0,1]^\Omega$ | Fuzzy power set of $\Omega$ |
| $A, B$ | Subsets of $\Omega$ |
| $A_j$ | Element of $\mathcal{F}$ |
| $\vec{A}$ | Measurement record |
| $C$ | Count of subset of $\Omega$ |
| $D$ | Data set |
| $E$ | Endpoints in possibilistic histogram |
| $F$ | Subset of $\Omega$ |
| $\tilde{F}, \tilde{G}$ | Fuzzy subsets of $\Omega$ |
| $\tilde{F}_\alpha$ | Alpha cut of fuzzy subset |
| $G_k$ | Domain invterval of possibilistic histogram |
| $M$ | $\lvert\vec{A}\rvert$, set of model states |
| $N$ | $\lvert\mathcal{S}\rvert$ |
| $O$ | Model |
| $P$ | Frequency measure, cumulative probability distribution |
| $Q$ | $\lvert E\rvert$ |
| $R$ | Fuzzy matrix |
| $R^{(j)}$ | $j$'th column of a fuzzy matrix |
| $S_i$ | Modeling or modeled system |
| $T_k$ | Function interval of possibilistic histogram |
| $U$ | Uniform random variable on $[0,1]$ |
| $V$ | Value set of semirings |
| $W$ | Range of order statistics, set of world states in a model |

| Symbol | Meaning |
|---|---|
| $X$ | Input alphabet of automaton |
| $Y$ | Output alphabet of automaton |
| $a$ | Atom of boolean algebra |
| $b$ | Bit; element of boolean algebra |
| $c$ | Count of singletons |
| $d$ | Data point |
| $e$ | Endpoint of possibilistic histogram |
| $f$ | Fit, frequency, model prediction function |
| $g$ | Structural aggregation function |
| $h$ | Numerical aggregation function |
| $i$ | Counter on $\Omega$ to $n$ |
| $j$ | Counter on $\mathcal{S}$ to $N$ |
| $k$ | Counter on $E$ to $Q$ |
| $l$ | Left interval endpoint |
| $m$ | Evidence function, states of a model |
| $n$ | $\|\Omega\|$ |
| $o$ | Observation function in a model |
| $p, \vec{p}$ | Probability distribution |
| $\vec{p}^*$ | Maximally uninformative (uniform) probability distribution |
| $q, \vec{q}$ | Generalized distribution |
| $q_\nu$ | Distribution of a fuzzy measure |
| $r$ | Right interval endpoint, "reality" function in a model |
| $s$ | Index of observations in $\vec{A}$ |
| $t$ | Time |
| $u$ | Possibilistic focus |
| $v$ | Hyperstate |
| $w$ | Sugeno's measure of fuzziness |
| $x, y, z$ | Dimensions of $\Omega$ in examples, elements of $\mathcal{L}$ |
| $\mathbf{C}$ | Core |
| $\mathbf{D}$ | Selected points for possibilistic histogram continuous approximation |
| $\mathbf{F}$ | Instrument ensemble |
| $\mathbf{H}$ | Entropy |
| $\mathbf{K}$ | Candidate points for possibilistic histogram continuous approximation |
| $\mathbf{N}$ | Nonspecificity |
| $\mathbf{P}$ | Conditional probability matrix |
| $\mathbf{S}$ | Strife |
| $\mathbf{T}$ | Total uncertainty |
| $\mathbf{U}$ | Support of fuzzy set or possibility distribution |
| $\mathbf{c}$ | Midpoint of possibilistic histogram core |
| $\mathbf{h}$ | Midpoints of $T_k$ |

| Symbol | Meaning |
|---|---|
| $\mathbf{l}$ | Left endpoint of support of $\pi$ at the origin |
| $\mathbf{r}$ | Right endpoint of support of $\pi$ at the origin |
| $\mathbf{t}^l, \mathbf{t}^r$ | Left and right endpoints of $T_k$ |
| $\mathcal{A}$ | Automata |
| $\mathcal{B}$ | Boolean algebra |
| $\mathcal{C}$ | Class on $\Omega$, general measuring device |
| $\mathcal{F}$ | Focal set |
| $\hat{\mathcal{F}}$ | Consistently transformed focal set |
| $\mathcal{L}$ | Lattice |
| $\mathcal{M}$ | Atoms of an algebra $\mathcal{B}$ |
| $\mathcal{R}$ | Semiring |
| $\mathcal{S}$ | Random set |
| $\hat{\mathcal{S}}$ | Consistently transformed random set |
| $\hat{\mathcal{S}}_i$ | Focused consistently transformed random set |
| $\mathcal{S}^\pi$ | Constructed consonant random set |
| $\mathcal{W}$ | Whole numbers |
| $\mathcal{Z}$ | Conorm process |
| $\mathcal{Z}^*$ | General process |
| $\mathbb{R}$ | Real numbers |
| Bel | Belief |
| Pl | Plausibility |
| $\vec{\text{Pl}}$ | Plausibility assignment |
| $\text{Pl}_i$ | Element of plausibility assignment |
| Pr | Probability |
| $\Gamma$ | Random set problem solution set |
| $\Gamma(\mathcal{S})$ | Set of focused, consistent transformations of $\mathcal{S}$ |
| $\Delta$ | Set of data intervals |
| $\Theta$ | Set of unitary column elements for row of $\boldsymbol{\Pi}$ |
| $\Lambda$ | Level set |
| $\Pi$ | Possibility measure |
| $\Pi^*$ | Possibility measure constructed from possibility distribution |
| $\boldsymbol{\Pi}$ | Conditional possibility matrix |
| $\Sigma$ | Sigma-field of $\Omega$ |
| $\Omega$ | Universe of discourse |
| $\alpha$ | $\in [0,1]$, for alpha-cuts |
| $\beta$ | Bound column in possibilistic process |
| $\gamma$ | Compatibility function |
| $\delta$ | Data interval, state transition function |
| $\eta$ | Necessity |
| $\theta$ | Unitary column element for row of $\boldsymbol{\Pi}$ |

| Symbol | Meaning |
|---|---|
| $\lambda$ | Automaton output function |
| $\mu$ | Membership function of |
| $\nu$ | Fuzzy measure |
| $\xi$ | Loss function |
| $\pi, \vec{\pi}$ | Possibility distribution |
| $\vec{\pi}^*$ | Maximally uninformative possibility distribution |
| $\sigma$ | State transition function |
| $\phi$ | State function |
| $\varphi$ | Complement operator |
| $\chi$ | Charactersitic function of |
| $\omega$ | Element of universe $\Omega$ |
| $\omega^*$ | Focus of a possibility distribution |
| $\sqcup$ | T-conorm operator |
| $\sqcap$ | T-norm operator |
| $\sqcup_m, \sqcap_m$ | Nilpotent conorm/norm operators |
| $\sqcup_w, \sqcap_w$ | Crisp conorm/norm operators |
| $\star$ | Monotonic conorm operator |
| $\vee$ | Maximum operator, lattice join |
| $\wedge$ | Minimum operator, lattice meet |
| $\oplus$ | Generalized disjunction/normalization operator |
| $\otimes$ | Generalized conjunction operator |
| $\circ$ | Fuzzy matrix composition |
| $\cdot$ | Ordinary matrix multiplaction |
| $\odot$ | Dempster combination of evidence functions |
| $\langle\cdot\rangle$ | Vector |
| $\tilde{\subseteq}$ | Fuzzy subset relation |
| $\preceq$ | Consonance ordering of disjoint devices |
| $\perp$ | Disjointness relation |
| $:=$ | Definition |
| max | Maximal optimization |
| min | Minimal optimization |
| $X \xrightarrow{\text{is-a}} Y$ | Class $Y$ inherits from class $X$ |
| $X \xrightarrow{\text{collection}} Y$ | Objects of class $Y$ are a collection of objects of class $X$ |
| $X \xrightarrow{\text{has-a}} Y$ | Objects of class $X$ logically determine objects of class $Y$ |

# References


[1] Aczel, J and Daroczy, Z: (1975) *On Measures of Information and Their Characterizations*, Academic Press, New York

[2] Akiyama, Y; Abe, T; and Mitsunaga, T et al.: (1991) "Conceptual Study of Max-* Composition on Correspondence of Base Space and its Applications in Determining Fuzzy Relations", *Japanese J. of Fuzzy Theory and Systems*, v. **3**:2, pp. 113-132

[3] Arbib, Michael A and Manes, Ernst: (1977) "Category Theoretic Approach to Systems in a Fuzzy World", in: *Systems: Approahces, Theories, Applications*, ed. WE Hartnett, pp. 1-26, D. Reidel

[4] Arbib, Michael A and Rhodes, JL et al.: (1968) "Complexity and Group Complexity of Finite State Machines and Finite Semi-Groups", in: *Algorithmic Theory of Machines, Languages, and Semi-Groups*, ed. MA Arbib, pp. 127-145, Academic Press, New York

[5] Artstein, Z and Vitale, RA: (1975) "Strong Law of Large Numbers for Random Compact Sets", *Annals of Probability*, v. **3**, pp. 879-882

[6] Ashby, Ross: (1956) *Introduction to Cybernetics*, Methuen, London

[7] Atkinson, M and Bancilhon, F, et al.: (1989) "Object-Oriented Database System Manifesto", in: *Proc. Deductive and Object-Oriented Databases*, Kyoto

[8] Barford, NC: (1967) *Experimental Measurements: Precision, Error and Truth*, John Wiley, New York

[9] Barone, Joseph M; Filev, Dimitar P; and Yager, Ronald R: (1993) "Exponential Possibility Distributions and the Mountain Method", in: *Proc. NAFIPS 1993*, ed. M. McAllister, pp. 118-122, Allentown PA

[10] Beltrami, Edward: (1987) *Mathematics for Dynamic Modeling*, Academic Press, Orlando

[11] Bennett, Charles H: (1988) "Dissipation, Information, Computational Complexity and the Definition of Organization", in: *Emerging Syntheses in Science*, ed. David Pines, pp. 215-234, Addison-Wesley, New York

[12] Bertalanffy, Ludwig von: (1968) *General Systems Theory*, George Braziller, New York

[13] Bezdek, James C; Hathaway, Richard J; and Sabin, MJ et al.: (1987) "Convergence Theorems for Fuzzy *c*-Means: Counterexamples and Repairs", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **17**:5, pp. 873-877

[14] Blockley, DI: (1979) "Role of Fuzzy Sets in Civil Engineering", *Fuzzy Sets and Systems*, v. **2**, pp. 267-278

[15] Blonsky, Marshall ed.: (1985) *On Signs*, John Hopkins U Press, Baltimore

[16] Bobrow, Daniel G ed.: (1985) *Qualitative Reasoning about Physical Systems*, MIT Press, Cambridge MA

[17] Bogler, P: (1987) "Shafer-Dempster Reasoning with Applications to Multitarget Identification of Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, pp. 968-977

[18] Bouchon, Bernadette: (1987) "Fuzzy Inferences and Conditional Possibility Distributions", *Fuzzy Sets and Systems*, v. **23**:2, pp. 33-41

[19] Brillouin, Leon: (1953) "Negentropy Principle of Information", *J. Applied Physics*, v. **24**:9, pp. 1152-1163

[20] Brooks, DR and Wiley, EO: (1988) *Evolution as Entropy*, U of Chicago Press, Chicago, 2nd edition

[21] Brumbaugh, David: (1992) *Object-Oriented Development: Building CASE Tools with C++*, John Wiley, NY

[22] Buckley, JJ: (1991) "Fuzzy Dynamical Systems I: The Discrete Case", in: *Proc. IFSA 1991*, v. **C**, pp. 16-18

[23] Cao, Hong Xing: (1987) "Fuzzy Verification of Weather Forecasts and Climate Simulation", in: *Analysis of Fuzzy Information*, v. **III**, ed. James Bezdek, CRC Press, Boca Raton

[24] Cao, Hong Xing and Chen, G: (1983) "Some Applications of Fuzzy Sets to Meteorological Forecasting", *Fuzzy Sets and Systems*, v. **9**, pp. 1-12

[25] Cariani, Peter A: (1991) "Epistemic Implications of Devices Which Construct Their Own Sensors and Effectors", in: *Proc. European Conf. on Artificial Life*

[26] Carlyle, JW: (1969) "Stochastic Finite-State System Theory", in: *System Theory*, ed. LA Zadeh and E Polak, pp. 387-423, MacGraw Hill

[27] Cattaneo, Gianpiero: (1991) "Fuzzy Operations and BZ Preclusivity Logics in Hilbert Space Quantum Physics", in: *Proc. IFSA 1991*, v. **C**, pp. 31-35

[28] Cellier, Francois: (1987) "Qualitative Simulation of Technical Systems Using the GSPS Framework", *Int. J. of General Systems*, v. **13**:4, pp. 333-344

[29] Chaitin, Gregory J: (1977) "Algorithmic Information Theory", *IBM J. Research Development*, v. **21**:4, pp. 350-359

[30] Chanas, Stefan and Heilpern, Stanislaw: (1989) "Single Value Simulations of Fuzzy Variables: Some Further Results", *Fuzzy Sets and Systems*, v. **33**, pp. 29-36

[31] Chanas, Stefan and Nowakowski, M: (1988) "Single Value Simulation of Fuzzy Variable", *Fuzzy Sets and Systems*, v. **25**, pp. 43-57

[32] Cheeseman, Peter: (1985) "In Defense of Probability", in: *Proc. Joint Conf. on Artificial Intelligence*, pp. 1002-1009, Los Angeles

[33] Cheeseman, Peter: (1986) "Probabilistic versus Fuzzy Reasoning", in: *Uncertainty in Artificial Intelligence*, ed. LN Kanal and JF Lemmer, pp. 85-102, North-Holland, Amsterdam

[34] Christensen, Ronald: (1980) *Entropy Minimax Sourcebook*, v. **1-4**, Entropy Limited, Lincoln, MA

[35] Christensen, Ronald: (1985) "Entropy Minimax Multvariate Statistical Modeling: I", *Int. J. of General Systems*, v. **11**, pp. 231-277

[36] Coad, Peter and Yourdon, Edward: (1991) *Object-Oriented Analysis*, Prentice-Hall, Eng.Cliffs NJ

[37] Codd, EF: (1968) *Cellular Automata*, Academic Press, New York

[38] Cooman, Gert de; Kerre, E; and Vanmassenhov, FR: (1992) "Possibility Theory: An Integral Theoretic Approach", *Fuzzy Sets and Systems*, v. **46**, pp. 287-299,

[39] Cox, Brad J: (11/90) "Planning the Software Industrial Revolution", *IEEE Software*, to appear

[40] Cox, Brad J: (1986) *Object Oriented Programming: An Evolutionary Approach*, Addison-Wesley, Reading MA

[41] Cumani, Aldo: (1982) "On a Possibilistic Approach to the Analysis of Fuzzy Feedback Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **12**:3, pp. 417-422,

[42] Dal Cin, M: (1975) "Fuzzy-State Automata: Stability and Fault Tolerance", *Int. J. Computer and Information Science*, v. **4**, pp. 63-80

[43] Dal Cin, M: (1975) "Modification Tolerance of Fuzzy-State Automata", *Int. J. Computer and Information Science*, v. **4**, pp. 81-93

[44] David, Herbert A: (1981) *Order Statistics*, Wiley, New York

[45] Davis, R and Hamscher, W: (1992) "Model-Based Reasoning: Troubleshooting", in: *Readings in Model-Based Diagnosis*, ed. W Hamscher et al., pp. 3-24, Morgan Kaufman, San Mateo CA

[46] Degroot, Morris H and Eddy, William F: (1983) "Set-Valued Parameters and Set-Valued Statistics", in: *Recent Advances in Statistics*, pp. 175-195, Academic Press

[47] Delgado, M and Moral, S: (1987) "On the Concept of Possibility-Probability Consistency", *Fuzzy Sets and Systems*, v. **21**, pp. 311-318

[48] Dempster, AP: (1967) "Upper and Lower Probabilities Induced by a Multivalued Mapping", *Annals of Mathematical Statistics*, v. **38**, pp. 325-339

[49] Doberkat, EE: (1978) *Stochastic Automata: Stability, Nondeterminism, and Prediction*, in: *Lecture Notes in Computer Science*, v. **11**, Springer-Verlag, Berlin

[50] Dougherty, E and Giardina, C: (1988) *Mathematical Methods for AI and Autonomous Systems*, Prentice-Hall

[51] Doyle, RJ; Sellers, SM; and Atkinson, DJ: (1989) "Focused Context-Sensitive Approach to Monitoring", in: *Proc. 11th Int. Joint Conf. on Artificial Intelligence*, Detroit

[52] Dubois, Didier: (1984) "Steps to a Theory of Qualitative Possibility", in: *Proc. 6th Int. Congress on Cybernetics and Systems*, pp. 10-14

[53] Dubois, Didier; Lang, Jèrôme; and Prade, Henri: (1992) "Dealing with Multi-Source Information in Possibilistic Logic", in: *Proc. 10th European Conf. on Artificial Intelligence*, pp. 38-42, Vienna

[54] Dubois, Didier and Prade, Henri: (1978) "Operations on Fuzzy Numbers", *Int. J. Systems Science*, v. **9**, pp. 613-626

[55] Dubois, Didier and Prade, Henri: (1980) *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York

[56] Dubois, Didier and Prade, Henri: (1982) "Class of Fuzzy Measures Based on Triangle Inequalities", *Int. J. of General Systems*, v. **8**, pp. 43-61

[57] Dubois, Didier and Prade, Henri: (1982) "On Several Representations of an Uncertain Body of Evidence", *Fuzzy Information and Decision Processes*, pp. 167-181

[58] Dubois, Didier and Prade, Henri: (1983) "Unfair Coins and Necessity Measures", *Fuzzy Sets and Systmes*, v. **10**, pp. 15-20

[59] Dubois, Didier and Prade, Henri: (1986) "Set Theoretic View of Belief Functions", *Int. J. of General Systems*, v. **12**:3, pp. 193-226

[60] Dubois, Didier and Prade, Henri: (1986) "Fuzzy Sets and Statistical Data", *European J. of Operational Research*, v. **25**, pp. 345-356

[61] Dubois, Didier and Prade, Henri: (1987) "Fuzzy Numbers: An Overview", in: *Analysis of Fuzzy Information*, v. **1**, pp. 3-39, CRC Press, Boca Raton

[62] Dubois, Didier and Prade, Henri: (1987) "Mean Value of a Fuzzy Number", *Fuzzy Sets and Systems*, v. **24**:3, pp. 279-300

[63] Dubois, Didier and Prade, Henri: (1987) "Principle of Minimum Specificity as a Basis for Evidential Reasoning", in: *Uncertainty in Knowledge-Based Systems*, ed. B. Bouchon and R. Yager, pp. 75-84, Springer-Verlag, Berlin

[64] Dubois, Didier and Prade, Henri: (1988) *Possibility Theory*, Plenum Press, New York

[65] Dubois, Didier and Prade, Henri: (1989) "Fuzzy Sets, Probability, and Measurement", *European J. of Operational Research*, v. **40**:2, pp. 135-154

[66] Dubois, Didier and Prade, Henri: (1989) "Measure-Free Conditioning, Probability, and Non-Monotonic Reasoning", in: *Proc. 11th Int. Joint Conf. on Artificial Intelligence*, pp. 1110-1114

[67] Dubois, Didier and Prade, Henri: (1990) "Logical View of Conditioning and its Application to Possibility and Evidence Theory", *Int. J. Approximate Reasoning*, v. **4**, pp. 23-46

[68] Dubois, Didier and Prade, Henri: (1990) "Consonant Approximations of Belief Functions", *Int. J. Approximate Reasoning*, v. **4**, pp. 419-449

[69] Dubois, Didier and Prade, Henri: (1990) "Modeling Uncertain and Vague Knowledge in Possibility and Evidence Theory", in: *Uncertainty in Artificial Intelligence 4*, ed. R. Schachter et al., pp. 307-318, North Holland

[70] Dubois, Didier and Prade, Henri: (1990) "Introduction to Possibilistic and Fuzzy Logics", in: *Readings in Uncertain Reasoning*, ed. G Shafer and J Pearl, pp. 742-761, Morgan-Kaufmana, Palo Alto CA

[71] Dubois, Didier; Prade, Henri; and Sandri, Sandra: (1991) "On Possibility/Probability Transformations", in: *Proc. IFSA 1991*, v. **M**, pp. 50-53

[72] Dubois, Didier and Prade, Henri: (1991) "Random Sets and Fuzzy Interval Analysis", *Fuzzy Sets and Systems*, v. **42**, pp. 87-101

[73] Dubois, Didier and Prade, Henri: (1991) "Inference in Possibilistic Hypergraphs", in: *Uncertainty in Knowledge Bases*, pp. 250-259

[74] Dubois, Didier and Prade, Henri: (1992) "Evidence, Knowledge and Belief Functions", *Int. J. Approximate Reasoning*, v. **6**:3, pp. 295-320

[75] Dubois, Didier and Prade, Henri: (1993) "Fuzzy Sets and Probability: Misunderstandings, Bridges and Gaps", in: *Proc. FUZZ-IEEE '93*

[76] Durrant-Whyt, HF: (1987) "Consistent Integration and Propogation of Disparate Sensor Observations", *Int. J. Robotics Research*, v. **6**:3, pp. 3-24

[77] Dvorak, D and Kuipers, B: (1992) "Model-Based Monitoring of Dynamic Systems", in: *Readings in Model-Based Diagnosis*, ed. W Hamscher et al., pp. 249-254, Morgan-Kaufmann, San Mateo CA

[78] Eco, Umberto: (1979) *Theory of Semiotics*, Indiana U Press, Bloomington

[79] Eco, Umberto: (1986) *Semiotics and the Philosophy of Language*, Indiana U Press, Bloomfield

[80] Eilenberg, S: (1974) *Automata, Languages and Semi-Groups*, v. **A**, Academic Press, New York

[81] Elias, Doug: (1986) "Systems Toolkit: A Collection of Systems Methods", in: *Proc. 30th Conf. of the Society for General Systems Research*, v. **1**

[82] Elsasser, Walter M: (1970) "Role of Individuality in Biological Theory", in: *Towards a Theoretical Biology*, v. **3**, ed. CH Waddington, pp. 153, Edinburgh U Press, Edinburgh

[83] Fine, TL: (1973) *Theories of Probability: An Examination of Foundations*, Academic Press, New York

[84] Fishwick, Paul A: (1989) "Qualitative Methodology in Simulation Model Engineering", *Simulation*, v. **52**:3, pp. 95-101

[85] Fishwick, Paul A: (1990) "Toward an Integrated Approach to Simulation Model Engineering", *Int. J. of General Systems*, v. **17**, pp. 1-19

[86] Fishwick, Paul A: (1991) "Introduction to Fuzzy Simulation", in: *Proc. EuroCAST '91*, pp. 14-18

[87] Fishwick, Paul A: (1991) "Fuzzy Simulation: Specifying and Identifying Qualitative Models", *Int. J. of General Systems*, v. **19**, pp. 295-316

[88] Fishwick, Paul A and Luker, Paul A eds.: (1990) *Qualitative Simulation Modeling and Analysis*, Springer-Verlag, New York

[89] Fulton, Steven L: (1990) "Introduction to Model-Based Reasoning", *AI Expert*, v. **1/90**, pp. 48-55,

[90] Fung, Robert M and Chong, Chee Yee: (1986) "Metaprobability and the Dempster-Shafer in Evidential Reasoning", in: *Uncertainty in Artificial Intelligence*, ed. LN Kanal and JF Lemmer, pp. 295-302, North-Holland, Amsterdam

[91] Gaines, Brian R and Kohout, Ladislav J: (1975) "Possible Automata", in: *Proc. Int. Symposium on Multi-Valued Logic*, pp. 183-196

[92] Gaines, Brian R and Kohout, Ladislav J: (1976) "Logic of Automata", *Int. J. of General Systems*, v. **2**:4, pp. 191-208

[93] Galway, L: (1986) "Workstation Based Environment for the Statistical Analysis of Set-Valued Data", in: *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, pp. 332-334

[94] Geer, James and Klir, George: (1991) "Discord in Possibility Theory", *Int. J. of General Systems*, v. **19**, pp. 119-132

[95] Geer, James and Klir, George: (1992) "Mathematical Analysis of Information Preserving Transformations Between Probabilistic and Possibilistic Formulations of Uncertainty", *Int. J. of General Systems*, v. **20**:2, pp. 143-176

[96] Gertler, Janos J and Anderson, Kenneth C: (1992) "Evidential Reasoning Extension to Quantitative Model-Based Failure Diagnosis", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **22**:2, pp. 275-289

[97] Giering, Edward W and Kandel, Abraham: (1983) "Application of Fuzzy Set Theory to Modeling Competition in Ecological Systems", *Fuzzy Sets and Systems*, v. **9**, pp. 103-127

[98] Giles, Robin: (1982) "Foundations for a Theory of Possibility", in: *Fuzzy Information and Decision Processes*, pp. 183-195, North-Holland

[99] Glorioso, Robert M and Osorio, FCC: (1980) *Engineering Intelligent Systems*, pp. 472, DEC, Bedford

[100] Goguen, JA: (1967) "L-fuzzy Sets", *J. Mathematical Analysis and Applications*, v. **18**, pp. 145-174,

[101] Goldberg, Adele and Robson, David: (1989) *Smalltalk-80*, Addison-Wesley, Reading MA

[102] Goodman, IR: (1984) "Some New Results Concerning Random Sets and Fuzzy Sets", *Information Sciences*, v. **34**, pp. 93-113

[103] Goodman, IR: (1993) "Conditional Event Algebra: A New Approach", in: *Proc. ISUMA 1993*, ed. BM Ayyub, pp. 55-62, IEEE Comp.Soc. Press, College Park MD

[104] Goodman, IR and Nguyen, HT: (1986) *Uncertainty Models for Knowledge-Based Systems*, North-Holland, Amsterdam

[105] Goodman, IR and Nguyen, HT: (1988) "Conditional Objects and the Modeling of Uncertainties", in: *Fuzzy Computing*, pp. 119-138, North-Holland, Amsterdam

[106] Grossing, Gerhard and Zeilinger, Anton: (1988) "Quantum Cellular Automata", *Complex Systems*, v. **2**, pp. 197-208

[107] Grossmann, Wilfried and Werthner, Hannes: (1993) "Stochastic Approach to Qualitative Simulation Using Markov Processes", in: *Proc. 1993 Int. Joint Conf. on Artificial Intelligence*

[108] Guariso, Giorgio; Rizzoli, Andrea; and Werthner, Hannes: (1992) "Identification of Model Structure via Qualitative Simulation", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **22**:5, pp. 1075-1086

[109] Gurocak, HB and Lazaro, A de Sam: (1992) "Fuzzy Logic Approach for Handling Imprecise Measurement in Robot Assembly", in: *Proc. 1992 FUZZ-IEEE Conference*, ed. Jim Bezdek, pp. 915-922, IEEE, San Diego

[110] Hackett, JK and Shah, M: (1990) "Multi-Sensor Fusion: A Perspective", in: *Proc. 1991 Int. Conf. on Robotics and Automata*, pp. 1324-1330, Cincinnati

[111] Hacking, I: (1975) "All Kinds of Possibility", *Philosophical Review*, v. **84**, pp. 321-337

[112] Hacking, I: (1975) *Emergence of Probability*, Cambridge U Press, Cambridge

[113] Hall, Gardiner A; Schuetzle, James; and La Vallee, D et al.: (1992) "Architectural Development of Real-Time Fault Diagnositc Systems Using Model-Based Reasoning", in: *Proc. 1992 Goddard Conf. on Space Applications of AI*, ed. Steve Rash, pp. 77-86, NASA Goddard, Greenbelt MD

[114] Hall, L; Szabo, S; and Kandel, A: (1986) "On Derivation of Memberships for Fuzzy Sets in Expert Systems", *Information Sciences*, v. **40**, pp. 39-52

[115] Halmos, PR: (1950) *Measure Theory*, van Nostrand, New York

[116] Hamscher, W; Console, Luca; and Kleer, Johan de eds.: (1992) *Readings in Model-Based Diagnosis*, Morgan-Kaufman

[117] Harnad, Steven: (1989) "Symbol Grounding Problem", in: *Proc. CNLS Conf. on Emergent Computation*

[118] Hisdal, E: (1978) "Conditional Possibilities: Independence and Noninteraction", *Fuzzy Sets and Systems*, v. **1**:4, pp. 283-297

[119] Hisdal, E: (1988) "Are Grades of Membership Probabilities?", *Fuzzy Sets and Systems*, v. **25**:3, pp. 325-348

[120] Hisdal, E: (1988) "Philosophical Issues Raised by Fuzzy Set Theory", *Fuzzy Sets and Systems*, v. **25**:3, pp. 349-356

[121] Holland, John: (1976) *Adaptation in Natural and Artificial Systems*, U. Michigan, Ann Arbor

[122] Hopcroft, John E and Ullman, Jeffery D: (1979) *Introduction to Automata Theory Languages and Computation*, Addison-Wesley, Reading MA

[123] Horgan, John: (1993) "Death of Proof", *Scientific American*, v. **269**:4, pp. 74-89

[124] Hughes, GE and Creswell, MJ: (1968) *Introduction to Modal Logic*, Methuen, London

[125] Jain, Ramesh: (1976) "Tolerance Analysis Using Fuzzy Sets", *Int. J. Systems Science*, v. **7**:12, pp. 1393-1401

[126] Janzen, Thomas E: (1993) "C++ Classes for Fuzzy Logic, Sets, and Associative Memories", *C User's Journal*, v. **11/93**, pp. 55-71

[127] Jaynes, ET: (1957) "Information Theory and Statistical Mechanics", *Physical Review*, v. **106**:10, pp. 620-630

[128] Joslyn, Cliff: (1991) "On the Semantics of Entropy Measures of Emergent Phenomena", *Cybernetics and Systems*, v. **22**:6, pp. 631-640

[129] Kampis, George: (1987) "Some Problems of System Descriptions I: Function", *Int. J. of General Systems*, v. **13**, pp. 143-156

[130] Kampis, George: (1991) *Self-Modifying Systems*, Pergamon, Oxford

[131] Kampè de, Fériet J: (1982) "Interpretation of Membership Functions of Fuzzy Sets in Terms of Plausibility and Belief", in: *Fuzzy Information and Decision Processes*, ed. MM Gupta and E Sanchez, pp. 93-98, North-Holland, Amsterdam

[132] Kandel, Abraham: (1978) "Fuzzy Statistics and Forecast Evaluation", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **8**, pp. 396-401

[133] Kandel, Abraham: (1980) "Fuzzy Dynamical Systems and the Nature of Their Solutions", in: *Fuzzy Sets*, ed. PP Wang and SK Chang, pp. 93-121, Plenum, New York

[134] Kandel, Abraham: (1986) *Fuzzy Mathematical Techniques with Applications*, Addison-Wesley

[135] Kandel, Abraham and Lee, A: (1979) *Fuzzy Switching and Automata*, Crane Russak, New York

[136] Kanerva, Penti: (1988) *Sparse Distributed Memory*, MIT Press, Cambridge

[137] Kapur, JN and Kesavan, HK: (1992) *Entropy Optimization Principles with Applications*, Academic Press, San Diego

[138] Kaufmann, A: (1975) *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York

[139] Kaufmann, A. and Gupta, M.M.: (1985) *Introduction to Fuzzy Arithmetic*, Reinhold, New York

[140] Kay, James J and Schneider, Eric: "Life as a Manifestation of the Second Law of Thermodynamics", *Int. J. Mathematical and Computer Modeling*

[141] Keene, Sonya E: (1989) *Object-Oriented Programming in Common LISP*, Addison-Wesley, Reading MA

[142] Kendall, DG: (1974) "Foundations of a Theory of Random Sets", in: *Stochastic Geometry*, ed. EF Harding and DG Kendall, pp. 322-376, Wiley, New York

[143] Kennes, Robert: (1992) "Computational Aspects of Möbius Transformations of Graphs", in: *IEEE Trans. on Systems, Man, and Cybernetics*, v. **22**:2, pp. 201-223,

[144] Klir, George: (1985) "Complexity: Some General Observations", *Systems Research*, v. **2**:2, pp. 131-140

[145] Klir, George: (1985) *Architecture of Systems Problem Solving*, Plenum, New York

[146] Klir, George: (1989) "Is There More to Uncertainty than Some Probability Theorists Believe?", *Int. J. of General Systems*, v. **15**:4, pp. 347-378

[147] Klir, George: (1989) "Probability-Possibility Conversion", *Proc. 3rd IFSA Congress*, pp. 408-411

[148] Klir, George: (1990) "Principle of Uncertainty and Information Invariance", *Int. J. of General Systems*, v. **17**:2-3, pp. 249-275

[149] Klir, George: (1991) *Facets of Systems Science*, Plenum, New York

[150] Klir, George: (1991) "Measures and Principles of Uncertainty and Information: Recent Developments", in: *Information Dynamics*, ed. H. Atmanspacher, pp. 1-14, Plenum Press, New York

[151] Klir, George: (1991) "Generalized Information Theory", *Fuzzy Sets and Systems*, v. **40**, pp. 127-142

[152] Klir, George: (1991) "Some Applications of the Principle of Uncertainty Invariance", in: *Proc. Int. Fuzzy Engineering Symposium*, pp. 15-26, Yakohama

[153] Klir, George: (1992) "Probabilistic vs. Possibilistic Conceptualization of Uncertainty", in: *Analysis and Management of Uncertainty*, ed. BM Ayyub et al., pp. 13-25, Elsevier

[154] Klir, George: (1993) "Developments in Uncertainty Based Information", in: *Advances in Computers*, v. **36**, ed. M. Yovitz, pp. 255-332, Academic Press

[155] Klir, George and Folger, Tina: (1987) *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall

[156] Klir, George and Geer, James: (1992) "Information Preserving Probability-Possibility Transformations: Recent Developments", in: *Fuzzy Logic: State of the Art*, ed. R. Lowen, Kluwer, to appear

[157] Klir, George and Mariano, M: (1987) "On the Uniqueness of Possibilistic Measure of Uncertainty and Information", *Fuzzy Sets and Systems*, v. **24**, pp. 197-219

[158] Klir, George and Parviz, Behvad: (1992) "Possibility-Probability Conversions: An Empirical Study", in: *Progress in Cybernetics and Systems*, ed. R. Trappl, pp. 19-26, Hemisphere, New York

[159] Klir, George and Parviz, Behvad: (1992) "On Probability-Possibility Transformations", in: *Proc. NAFIPS 1992*, v. **2**, pp. 598-607, Puerto Vallarta

[160] Klir, George and Parviz, Behvad: (1992) "Probability-Possibility Transformations: A Comparison", *Int. J. of General Systems*, v. **21**:1, pp. 291-310

[161] Kloeden, PE: (1982) "Fuzzy Dynamic Systems", *Fuzzy Sets and Systems*, v. **7**, pp. 275-296

[162] Kolmogorov, AN: (1965) "Three Approaches to the Quantitative Definition of Information", *Problems of Information Transmission*, v. **1**:1, pp. 1-7

[163] Korson, Tim and McGregor, John D: (1990) "Understanding Object-Oriented: A Unifying Paradigm", *Communications of the ACM*, v. **33**:9, pp. 40-60

[164] Kosko, B: (1990) "Fuzziness vs. Probability", *Int. J. of General Systems*, v. **17**:2-3, pp. 211-240

[165] Kosko, B ed.: (1990) *Neural Networks and Fuzzy Systems*, Prentice-Hall, Eng. Cliffs NJ

[166] Kovalerchuk, B and Taliansky, V: (1992) "Comparison of Empirical and Computed Values of Fuzzy Conjunction", *Fuzzy Sets and Systems*, v. **46**, pp. 49-53

[167] Krantz, DH; Luce, RD; and Suppes, P, et al.: (1971) *Foundations of Measurement*, Academic Press, New York

[168] Kreinovich, V; Quintana, C; and Reznik, L: (1992) "Gaussian Membership Functions are the Most Adequate in Representing Uncertainty in Measures", in: *Proc. NAFIPS 1992*, v. **2**, pp. 618-624, NASA, Puerto Vallerta

[169] Kripke, Saul: (1963) "Semantical Analysis of Modal Logic I: Normal Propositional Calculus", *Zeitschrift für Mathematische*, v. **9**, pp. 67-96

[170] Krishnapuram, Raghu and Keller, James: (1993) "Possibilistic Approach to Clustering", *IEEE Trans. on Fuzzy Systems*, v. **1**:2, pp. 98-110

[171] Kuipers, BJ: (1989) "Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge", *Automatica*, v. **25**, pp. 571-585

[172] Langton, Chris ed.: (1988) *Artificial Life*, Addison-Wesley

[173] Lasker, George E: (1981) *Fuzzy Sets and Systems, Possibility Theory and Special Topics*, in: *Applied Systems and Cybernetics*, v. **VI**, Pergamon, London

[174] Lee, CC: (1990) "Fuzzy Logic in Control Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **20**:2, pp. 404-435

[175] Lee, ET and Zadeh, LA: (1969) "Note on Fuzzy Languages", *Information Sciences*, v. **1**, pp. 421,

[176] Lemmer, JF: (1986) "Confidence Factors, Empiricism, and the Dempster-Shafer Theory of Evidence", in: *Uncertainty in Artificial Intelligence*, ed. LN Kanal and JF Lemmer, pp. 117-125, North-Holland, Amsterdam

[177] Lindenmayer, A: (1968) "Mathematical Models for Cellular Interaction in Development", *J. Theoretical Biology*, v. **18**, pp. 280-315

[178] Looney, Carl G: (1988) "Fuzzy Petri Nets for Rule-Based Decisionmaking", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **18**:1, pp. 178-183

[179] Lorenz, EN: (1963) "Deterministic Nonperiodic Flow", *J. Atmospheric Sciences*, v. **20**

[180] Luo, RC and Kay, MG: (1989) "Multisensor Integration and Fusion in Intelligent Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **19**, pp. 901-931

[181] Mahler, Ronald: (1993) "Random Sets as a Foundation for General Data Fusion", in: *Proc. Joint Service Data Fusion Symposium*, Orlando, FL

[182] Matheron, G: (1975) *Random Sets and Integral Geometry*, Wiley, New York

[183] McGoveran, David: (1980) "Fuzzy Logic and Non-Distributive Truth Valuations", in: *Fuzzy Sets*, ed. PP Wang and SK Chang, pp. 49-57, Plenum, New York

[184] Mittelmann, Rudolf: (1988) "Object Oriented Implementation of Petri Nets Concepts", in: *Cybernetics and Systems '88*, ed. R. Trappl, pp. 759-766, Kluwer, Boston

[185] Mittelstaedt, P; Prieur, A; and Schieder, R: (1987) "Unsharp Duality in a Two Doubleslit Experiment", *Foundations of Physics*, v. **17**:9, pp. 891-903

[186] Mizumota, Masaharu; Toyoda, Junichi; and Tanaka, Kohkichi: (1969) "Some Considerations of Fuzzy Automata", *J. Computer and Systems Science*, v. **3**, pp. 409-422

[187] Moore, RM: (1979) *Methods and Applications of Interval Analysis*, in: *SIAM Studies in Applied Mathematics*, SIAM, Philadelphia

[188] Morris, Charles: (1938) *Foundations of the Theory of Signs*, v. **1,2**, U. Chicago, Chicago

[189] Močkor, Jiři: (1991) "Category of Fuzzy Automata", *Int. J. of General Systems*, v. **20**, pp. 73-82

[190] Natvig, B: (1983) "Possibility versus Probability", *Fuzzy Sets and Systems*, v. **10**, pp. 31-36

[191] Nguyen, HT: (1978) "On Conditional Possibility Distributions", *Fuzzy Sets and Systems*, v. **1**:4

[192] Nguyen, HT: (1978) "On Random Sets and Belief Functions", *J. Mathematical Analysis and Applications*, v. **65**, pp. 531-542

[193] Näther, Wolfgang: (1990) "On Possibilistic Inference", *Fuzzy Sets and Systems*, v. **36**, pp. 327-337

[194] Orchard, Robert: (1990) "Knowledge Processing: Semantics for the Klir Hierarchy of General Systems", in: *Computer Aided Systems Theory*, ed. Franz Pichler, pp. 33-40, Springer-Verlag, Berlin

[195] Ören, Tuncer I: "GEST—A Modeling and Simulation Language Based on Systems Theoretic Concepts", in: *Simulation and Model-Based Methodologies*, ed. TI Ören, BP Zeigler, pp. 281-335, Springer-Verlag, Heidelberg

[196] Ören, Tuncer I: (1989) "Bases for Advanced Simulation", in: *Modeling and Simulation Methodology*, ed. MS Elzas, TI Ören, pp. 29-43, North-Holland, Amsterdam

[197] Otto, Kevin N and Antonsson, Erik: (1991) "Trade-Off Strategies in Engineering Design", *Research in Engineering Design*, v. **3**, pp. 87-103

[198] Park, Daihee; Kandel, Abraham; and Langholz, Gideon: (1994) "Genetic-Based New Fuzzy Reasoning Models with Application to Fuzzy Control", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **24**:1, pp. 39-47

[199] Pattee, Howard: (1982) "Cell Psychology: An Evolutionary View of the Symbol-Matter Problem", *Cognition and Brain Theory*, v. **5**, pp. 325-341

[200] Pattee, Howard: (1985) "Universal Principle of Measurement and Language Function in Evolving Systems", in: *Complexity, Language, and Life*, ed. John Casti, pp. 268-281, Springer-Verlag, Berlin

[201] Pattee, Howard: (1988) "Simulations, Realizations, and Theories of Life", in: *Artificial Life*, ed. C Langton, pp. 63-77, Addison-Wesley, Redwood City CA

[202] Pattee, Howard: (1990) "Measurement Problem in Physics, Computation, and Brain Theory", in: *Nature, Cognition and System*, v. **2**, ed. ME Carvallo, Kluwer, Netherlands

[203] Paz, A: (1971) *Introduction to Probabilistic Automata*, Academic, New York

[204] Pedrycz, W: (1981) "Approach to the Analysis of Fuzzy Systems", *Int. J. Control*, v. **34**, pp. 403-421

[205] Peeva, Kety: (1991) "Equivalence, Reduction and Minimization of Finite Automata over Semirings", *Theoretical Computer Science*, v. **88**:2, pp. 269-285

[206] Peirce, Charles S. ed.: (1955) *Philosophical Writings of Peirce*, Dover

[207] Person, Michael: (1991) "Fuzzy Petri Net Languages", in: *Proc. IFSA 1991*, v. **C**, pp. 201-204

[208] Peterson, JL: (1977) "Petri Nets", *ACM Computing Surveys*, v. **9**:3, pp. 223-252

[209] Petley, BW: (1988) *Fundamental Physical Constants and the Frontier of Measurement*, Adam Hilger, Bristol

[210] Pichler, Franz: (1988) "CAST – Computer Aided Systems Theory", in: *Cybernetics and Systems '88*, ed. R. Trappl, pp. 731-736, Kluwer, Boston

[211] Pichler, Franz ed.: (1990) *Computer Aided Systems Theory*, Springer-Verlag, Berlin

[212] Pichler, Franz and Moreno-Diaz, R eds.: (1990) *Computer Aided Systems Theory: EUROCAST 89*, in: *Lecture Notes in Computer Science*, v. **410**, Springer-Verlag, Berlin

[213] Pichler, Franz and Praehofer, Herbert: (1988) "CAST:FSM – Computer Aided Systems Theory: Finite State Machine", in: *Cybernetics and Systems '88*, ed. R. Trappl, pp. 737-742, Kluwer, Boston

[214] Pichler, Franz and Schwärtzel, H: (1992) CAST *Methods in Modeling*, Springer-Verlag, Berlin

[215] Powers, WT: (1973) *Behavior, the Control of Perception*, Aldine, Chicago

[216] Powers, WT ed.: (1989) *Living Control Systems*, CSG Press

[217] Prade, Henri: (1982) "Modal Semantics and Fuzzy Set Theory", in: *Fuzzy Sets and Possibility Theory*, ed. R Yager, pp. 232-246, Pergamon, Oxford

[218] Praehofer, Herbert: (1988) "STIDEVS: Systems Theory Instrumented Discrete Event Simulation", in: *Cybernetics and Systems '88*, ed. R. Trappl, pp. 767-773, Kluwer, Boston

[219] Praehofer, Herbert: (1991) "Systems Theory Instrumented Modeling and Simulation Methodology", *Cybernetics and Systems*, v. **22**, pp. 283-312

[220] Praehofer, Herbert: (1991) "Systems Theory Forms for Combined Discrete-Continuous Systems Simulation", *Int. J. of General Systems*, v. **19**, pp. 219-240

[221] Prigogine, Ilya: (1980) *From Being to Becoming*, WH Freeman, San Francisco

[222] Prigogine, Ilya and Nicolis, Gregoire: (1972) "Thermodynamics of Evolution", *Physics Today*, v. **25**, pp. 23-28

[223] Pulmannova, Sylvia: (1987) "Representation of Quantum Logics and Transition Probability Spaces", in: *Concept of Probability*, ed. EI Bitsakis et. al., pp. 51-59, Kluwer,

[224] Puri, ML and Ralescu, DA: (1986) "Fuzzy Random Variables", *J. Mathematical Analysis and Applications*, v. **114**, pp. 409-422

[225] Pykacz, Jarosław: (1987) "On Bell-Type Inequalities in Quantum Logics", in: *Concept of Probability*, ed. EI Bitsakis et al., pp. 115-120, Kluwer

[226] Pykacz, Jarosław: (1991) "Fuzzy Set Ideas in Foundations of Quantum Mechanics", in: *Proc. IFSA 1991*, v. **C**, pp. 205-208

[227] Ramer, Arthur: (1987) "Uniqueness of the Information Measure in the Theory of Evidence", *Fuzzy Sets and Systems*, v. **24**, pp. 183-196

[228] Ramer, Arthur: (1989) "Conditional Possibility Measures", *Cybernetics and Systems*, v. **20**, pp. 233-247

[229] Ramer, Arthur: (1990) "Axioms of Uncertainty Measures: Dependence and Independence", *Fuzzy Sets and Systems*, v. **35**, pp. 185-196

[230] Ramer, Arthur: (1990) "Concepts of Fuzzy Information Measures on Continuous Domains", *Int. J. of General Systems*, v. **17**:2-3, pp. 241-248

[231] Ramer, Arthur and Lander, Leslie: (1987) "Classification of Possibilistic Uncertainty and Information Functions", *Fuzzy Sets and Systems*, v. **24**, pp. 221-230

[232] Ramer, Arthur and Puflea-Ramer, Rodica: (1993) "Uncertainty as the Basis of Possibility Conditioning", in: *Proc. ISUMA 1993*, ed. BM Ayyub, pp. 79-82, IEEE Comp.Soc. Press, College Park MD

[233] Rao, MB and Rashed, A: (1981) "Some Comments on Fuzzy Variables", *Fuzzy Sets and Systems*, v. **6**, pp. 285

[234] Ray, AK and Chatterjee, B: (1991) "Formal Power Series Approach to the Construction of Minimal Fuzzy Automata", *Information Sciences*, v. **55**, pp. 189-207

[235] Reichanbach, H.: (1949) *Theory of Probability*, Berkeley

[236] Reichanbach, H.: (1991) *Direction of Time*, U California Press, Berkley

[237] Rescher, Nicholas: (1970) *Scientific Explanation*, The Free Press, New York

[238] Rescher, Nicholas: (1984) *Limits of Science*, U California Press, Berkeley

[239] Resconi, Germano; Klir, George J; and St. Clair, Ute: (1992) "Hierarchical Uncertainty Metatheory Based On Modal Logic", *Int. J. of General Systems*, v. **21**, pp. 23-50

[240] Roberts, DW: (1989) "Analysis of Forest Succession with Fuzzy Graph Theory", *Ecological Modeling*, v. **45**, pp. 261-274

[241] Rosen, Robert: (1985) *Anticipatory Systems*, Pergamon, Oxford

[242] Rosen, Robert: (1991) *Life Itself*, Columbia U Press, New York

[243] Rumelhart, D: (1986) *Parallel Distributed Processing, Foundations*, v. **1**, MIT Press, Cambridge, MA

[244] Ruspini, Enrique H: (1989) "Semantics of Vague Knowledge", *Revue Internationale de Systemique*, v. **3**:4, pp. 387-420

[245] Salmon, Welsey C.: (1970) "Statistical Explanation", in: *Nature and Function of Scientific Theories*, ed. RG Colodny, U. Pittsburg, Pittsburg

[246] Santos, E: (1968) "Maximin Automata", *Information and Control*, v. **13**, pp. 363-377

[247] Santos, E: (1970) "Fuzzy Algorithms", *Information and Control*, v. **17**, pp. 326-339

[248] Santos, E: (1972) "On Reductions of Maximin Machines", *J. Mathematical Analysis and Applications*, v. **40**, pp. 60-78

[249] Santos, E: (1972) "Max-Product Machines", *J. Mathematical Analysis and Applications*, v. **37**, pp. 677-686

[250] Santos, E: (1975) "Max-Product Grammars and Languages", *Information Sciences*, pp. 1-23

[251] Santos, E: (1977) "Fuzzy and Probabilistic Programs", in: *Fuzzy Automata and Decision Processes*, ed. MM Gupta, pp. 133-147

[252] Santos, E and Wee, WG: (1968) "General Formulation of Sequential Machines", *Information and Control*, v. **12**, pp. 5-10

[253] Sarna, M and Wojnarowski, J: (1989) "Fuzzy Models in Mechanics", *Strojnicky Časopis*, v. **40**:3, pp. 309-317

[254] Sassure, Ferdinand: (1959) *Course in General Linguistics*, McGraw-Hill, New York

[255] Schneider, Thomas D: (1991) "Theory of Molecular Machines I: Channel Capactiy", *J. Theoretical Biology*, v. **148**:1, pp. 125-137

[256] Schneider, Thomas D: (1991) "Theory of Molecular Machines II: Energy Dissipation", *J. Theoretical Biology*, v. **148**:1, pp. 83-123

[257] Schrödinge, E: (1967) *What is Life? and Other Scientific Essays*, Cambridge U Press, Cambridge

[258] Schuster, Heinz G: (1984) *Deterministic Chaos*, Physik-Verlag, Weinheim FRG

[259] Seidenfeld, Teddy: (1979) "Why I Am Not an Objective Bayesian", *Theory and Decision*, v. **11**, pp. 413-440

[260] Shackle, GLS: (1969) *Decision, Order and Time in Human Affairs*, Cambridge U Press, Cambridge UK

[261] Shafer, Glen: (1976) *Mathematical Theory of Evidence*, Princeton U Press, Princeton

[262] Shafer, Glen: (1978) "Non-Additive Probabilities in the Works Bernoulli and Lambert", *Arch. Hist. Exact Science*, v. **19**, pp. 309-370

[263] Shannon, CE: (1956) "Bandwagon", *IRE Trans. on Information Theory*, v. **2**:1, pp. 2-3

[264] Shannon, CE and Weaver, W: (1964) *Mathematical Theory of Communication*, U Illinois Press, Urbana

[265] Shen, Qiang and Leitch, Roy: (1991) "Diagnosing Continuous Dynamic Systems By the Use of Qualitative Simulation", in: *Proc. Third International Conf. on Control*, v. **2**, pp. 1000-1006, Edinburgh

[266] Shen, Qiang and Leitch, Roy: (1993) "Fuzzy Qualitative Simulation", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **23**:4, pp. 1038-1061

[267] Singer, Dennis: (1992) "Fuzzy-Set Approach to Non-Equilibrium Thermodynamics", *Fuzzy Sets and Systems*, v. **47**, pp. 39-48

[268] Singh, MG ed.: (1987) *Systems and Control Encyclopedia*, Pergamon, Oxford

[269] Skilling, John ed.: (1989) *Maximum-Entropy and Bayesian Methods*, Kluwer

[270] Smets, Philippe: (1990) "Constructing the Pignistic Probability Functionin a Context of Uncertainty", in: *Uncertainty in Artificial Intelligence 5*, ed. M Henrion et al., pp. 29-39, North Holland, New York

[271] Smithson, Michael: (1988) *Ignorance and Uncertainty: Emerging Paradigms*, Springer-Verlag, New York

[272] Sosnowski, Zenon A and Pedrycz, W: (1992) "FLISP: Representing and Processing of Uncertain Information", in: *Fuzzy Logic for the Management of Uncertainty*, ed. L Zadeh and J Kacpryzk, pp. 495-521, Wiley, New York

[273] Starke, PH: (1972) *Abstract Automata*, North Holland, Amsterdam

[274] Stroustrup, Bjarne: (1991) *C++ Programming Language*, Addison Wesley, 2nd edition

[275] Sudkamp, Thomas: (1992) "Semantics of Plausibility and Possibility", *Int. J. of General Systems*, v. **21**, pp. 273-289

[276] Sugeno, Michio: (1974) *Theory of Fuzzy Integrals and its Applications*, Tokyo Inst. Tech., Tokyo, PhD thesis

[277] Sugeno, Michio: (1977) "Fuzzy Measures and Fuzzy Integrals: A Survey", in: *Fuzzy Automata and Decision Processes*, ed. MM Gupta et al., pp. 89-102, North-Holland, New York,

[278] Sugeno, Michio and Yasukawa, Takahiro: (1993) "Fuzzy-Logic-Based Approach to Qualitative Modeling", *IEEE Trans. on Fuzzy Systems*, v. **1**:1, pp. 7-31

[279] Szilard, L: (1964) "On the Decrease of Entropy in Thermodynamic Systems by the Intervention of Intelligent Beings", *Behavioral Science*, v. **9**

[280] Tanaka, Hideo and Ishibuchi, Hisao: (1993) "Evidence Theory of Exponential Possibility Distribution", *Int. J. Approximate Reasoning*, v. **8**, pp. 123-140

[281] Tanir, Oryal and Sevnic, Suleyman: (1994) "Defining Requirements for a Standard Simulation Environment", *IEEE Computer*, v. **27**:2, pp. 28-34

[282] Taylor, Barry N and Kuyatt, Chris E: (1993) *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, NIST, Gaithersburg MD

[283] Terano, Toshiro; Asai, Kiyoji; and Sugeno, Michio: (1987) *Fuzzy Systems Theory and Its Applications*, Academic Press, New York

[284] Thoma, H Mathis: (1991) "Belief Function Computations", in: *Conditional Logic in Expert Systems*, ed. IR Goodman et al., pp. 269-308, North Holland

[285] Thomason, MG: (1973) "Finite Fuzzy Automata, Regular Fuzzy Languages and Pattern Recognition", *J. Pattern Recognition*, v. **5**, pp. 383-390

[286] Thomason, MG: (1977) "Convergence of Powers of a Fuzzy Matrix", *J. Mathematical Analysis and Applications*, v. **57**, pp. 476-480

[287] Toffoli, T and Margolus, N: (1986) *Cellular Atuomata Machines: A New Environment for Modeling*, MIT Press

[288] Topencharov, V and Peeva, K: (1981) "Equivalence, Reduction and Minimization of Finite Fuzzy Automata", *J. Mathematical Analysis and Applications*, v. **84**, pp. 270-281

[289] Tribus, Myron and Levine, RD eds.: (1979) *Maximum Entropy Formalism*, MIT Press, Cambridge

[290] Turchin, Valentin: (1977) *Phenomenon of Science*, Columbia U Press, New York

[291] Türksen, IB: (1984) "Measurement of Fuzziness: Interpretation of Axioms of Measurement", in: *Proc. Conf. on Fuzzy Information for Knowledge Representation and Decision Analysis*, pp. 97-102, IFAC, Oxford

[292] Ulanowicz, Robert E: (1986) *Growth and Development*, Ecosystems

[293] Voorbaak, Frans: (1989) "Computationally Efficient Approximations of Dempster-Shafer Theory", *Int. J. Man-Machine Studies*, v. **30**:5, pp. 525-536

[294] Wang, PP ed.: (1983) *Advances in Fuzzy Sets, Possibility Theory, and Applications*, Plenum, New York

[295] Wang, PZ: (1991) "Fuzziness vs. Randomness: Falling Shadow Theory", *BUSEFAL*, v. **48**, pp. 64-73

[296] Wang, PZ and Liu, XH: (1984) "Set-Valued Statistics", *J. Engineering Mathematics*, v. **1**:1, pp. 43-54

[297] Wang, PZ; Liu, XH; and Sanchez, E: (1986) "Set-Valued Statistics and its Application to Earthquake Engineering", *Fuzzy Sets and Systems*, v. **18**, pp. 347-356

[298] Wang, Zhenyuan: (1991) *Fuzzy Measure Theory*, SUNY-Binghamton, Binghamton NY, PhD disssertation

[299] Wang, Zhenyuan and Klir, George J: (1992) *Fuzzy Measure Theory*, Plenum Press, New York

[300] Weaver, Warren: (1968) "Science and Complexity", *American Scientist*, v. **36**, pp. 536-544

[WeBDeF88] Weber, Bruce; Depew, FJ; and Smith, James D eds.: (1988) *Entropy, Information, and Evolution*, MIT Press, Cambridge

[301] Webster, Miriam Co.: (1989) *New Merriam Webster Dictionary*, Merriam-Webster, Springfield MA

[302] Wee, WG and Fu, KS: (1969) "Formulation of Fuzzy Automata and its Application as a Model of a Learning System", *IEEE Trans. on Systems Science and Cybernetics*, v. **5**, pp. 215-223

[303] Wheeler, John A: (1990) "Information, Physics, Quantum: The Search for Links", in: *Complexity, Entropy and the Physics of Information*, pp. 3-28, Addison-Wesley, Reading MA

[304] Wierman, Mark J: (1991) "Possibilistic Image Processing", in: *Tenth SPIE Conf. on Algorithmic and Robot Vision*

[305] Wildberger, A Martin: (1993) "Induction of Knowledge From Data By Clustering", in: *Proc. 1993 Summer Simulation Conf.*, Soc. Computer Sim., Boston

[306] Windham, Michael P: (1982) "Cluster Validity for the Fuzzy $c$-Means Clustering Algorithm", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. **4**:4, pp. 357-363

[307] Wolf, Alan: (1986) "Quantifying Chaos with Lyapunov Exponents", *Nonlinear Science: Theory and Applications*, ed. Holden, pp. 273-290, Manchester U Press

[308] Wong, GA and Shen, DWC: (1972) "On the Learning Behavior of Fuzzy Automata", *Advances in Cybernetics and Systems*, v. **2**, pp. 885

[309] Wood, Kristin L and Antonsson, Erik K: (1989) "Computations with Imprecise Parameters in Engineering Design", *J. Mechanics, Transmissions and Automation*, v. **111**, pp. 616-625

[310] Wood, Kristin L; Otto, KN; and Antonson, EK: (1989) "Formal Method for Representing Uncertainties in Engineering Design", in: *First Int. Workshop on Formal Methods in Engineering Design*, Colorada Spring

[311] Yager, Ronald R: (1980) "Foundation for a Theory of Possibility", *J. Cybernetics*, v. **10**, pp. 177-204

[312] Yager, Ronald R: (1980) "On the Lack of Inverses in Fuzzy Arithmetic", *Fuzzy Sets and Systems*, v. **4**, pp. 73-

[313] Yager, Ronald R: (1981) "Some Properties of Fuzzy Relations", *Cybernetics and Systems*, v. **12**, pp. 123-140

[314] Yager, Ronald R ed.: (1982) *Fuzzy Set and Possibility Theory: Recent Development*, Pergamon, Oxford

[315] Yager, Ronald R: (1982) "Level Sets for Membership Evaluation of Fuzzy Subsets", in: *Fuzzy Sets and Possibility Theory: Recent Developments*, pp. 90-97, Pergamon, Oxford,

[316] Yager, Ronald R: (1986) "Paths of Least Resistance in Possibilistic Production Systems", *Fuzzy Sets and Systems*, v. **19**:2, pp. 121-132

[317] Yager, Ronald R: (1993) "Element Selection from a Fuzzy Subset Using Fuzzy Integral", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **23**:2, pp. 467-477

[318] Yager, Ronald R: (1993) "On the Completion of Qualitative Possibility Measures", *IEEE Trans. on Fuzzy Systems*, v. **1**:3, pp. 184-194

[319] Yeh, RT and Bang, SY: (1975) "Fuzzy Relations, Fuzzy Graphs and their Applications", in: *Fuzzy Sets and Applications to Cognitive and Decision Processes*, pp. 125-149, Academic Press, New York

[320] Zadeh, Lotfi A: (1965) "Fuzzy Sets and Systems", in: *Systems Theory*, ed. J. Fox, pp. 29-37, Polytechnic Press, Brooklyn NY

[321] Zadeh, Lotfi A: (1965) "Fuzzy Sets", *Information and Control*, v. **8**, pp. 338-353

[322] Zadeh, Lotfi A: (1973) "Outline of a New Approach to the Analysis of Complex Systems", *IEEE Trans. on Systems, Man, and Cybernetics*, v. **3**

[323] Zadeh, Lotfi A: (1975) "Calculus of Fuzzy Restrictions", in: *Fuzzy Sets and Applications to Cognitive and Decision Processes*, ed. LA Zadeh et. al., pp. 1-39, Academic Press, New York

[324] Zadeh, Lotfi A: (1975) "Concept of a Linguistic Variable and Its Application to Approximate Reasoning", *Information Sciences*, v. **8**, pp. 199-249

[325] Zadeh, Lotfi A: (1978) "Fuzzy Sets as the Basis for a Theory of Possibility", *Fuzzy Sets and Systems*, v. **1**, pp. 3-28

[326] Zeigler, BP: (1976) *Theory of Modeling and Simulation*, Wiley, New York

[327] Zeigler, BP: (1990) *Object-Oriented Simulation with Hierarchical Modular Models*, Academic Press, San Diego

[328] Zimmerman, HJ: (1991) *Fuzzy Set Theory*, Kluwer, Boston, 2nd edition

[329] Zurek, Wojciech ed.: (1990) *Complexity, Entropy, and the Physics of Information*, Addison-Wesley, Reading MA