Joslyn, Cliff A: (1996) "Hybrid Methods to Represent Incomplete and Uncertain Information", in: Proc. 1996 Interdisciplinary Conf. on Intelligent Systems: A Semiotic Perspective, ed. J. Albus, A. Meystel et al., pp. 133-140, NIST, Gaithersburg MD

Hybrid Methods to Represent Incomplete and Uncertain Information*

Cliff Joslyn †

July, 1996

Abstract

Decision making is cast in the semiotic context of perception, decision, and action loops. Towards the goal of properly grounding hybrid representations of information and uncertainty from this semiotic perspective, we consider the roles of and relations among the mathematical components of General Information Theory (GIT), particularly among fuzzy sets, possibility theory, probability theory, and random sets. We do so by using a clear distinction between the syntactic, mathematical formalism and the semantic domains of application of each of these fields, placing the emphasis on available measurement and action methods appropriate for each formalism, to which and from which the decision-making process flows.

1 Introduction

From a semiotic perspective [3, 17], an intelligent system can be described as a control system, involving a perception or measurement (input) function from the environment to the system, an action (output) function from the system to the environment, and a knowledge or information (throughput) function within the system. Another important consequence of the semiotic perspective is the clear distinction drawn among the syntactic, semantic, and pragmatic aspects of the use of information in systems. All of these are mutually necessary at different levels of analysis in a fully functioning, viable semiotic system.

We can describe decision making generally as the process of selecting among a variety of choices, whether completely (selecting a particular choice) or incompletely (reducing or constraining the viable choices). Thus decision making cannot be separated from the concepts of information and uncertainty: uncertainty is present whenever there are multiple possible choices; and the gain of information is represented by selecting among the choices, thus reducing that uncertainty.

Classically, probability theory has been the sole mathematical method to represent information and uncertainty; that is, probability distributions have been the only method by which weights are placed on choices. But recent years have seen a proliferation of new, non-probabilistic mathematical methods for the representation of uncertainty and information in systems models. Following Klir [14] we call these methods collectively "General Information Theory" (GIT), which includes fuzzy sets, systems, and logic [15]; fuzzy measures [21]; random set [13] and Dempster-Shafer evidence theory [6]; possibility theory [2]; imprecise probabilities [20]; probability bounds [5]; rough set theory [19]; and others.

Each of these involves some form of generalization or extension away from stochastic representations, and any or all of them can be utilized in semiotic systems to represent the information of perceptions, actions, and decisions. Transformations can be made among them, and hybrid representations can be utilized.

In this paper we first make the decision making process in semiotic systems more explicit, highlighting both the semantic and informational aspects. We then briefly introduce the components of GIT, and elucidate the relations among fuzzy sets, possibility theory, probability theory, and random sets. Finally, we consider more specifically the realization of these representations in semiotic systems, and their consequences for decision making in particular.

2 Semiotic Control Systems

Fig. 1 shows a system in a classical control relation with its environment. In the world (the system's environment) the processes of "reality" proceed outside of the knowledge of the system. Rather, all knowledge

*To be presented in the session on "Information Representation in Decision Making Systems" at the conference *Intelligent Systems: A Semiotic Perspective*, National Institute of Standards and Technology, October, 1996, `http://gwis2.circ.gwu.edu/~joslyn/sem96`.

†NRC Research Associate, Mail Code 522.3, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, `joslyn@kong.gsfc.nasa.gov`, `http://gwis2.circ.gwu.edu/~joslyn`, (301) 286-5773.

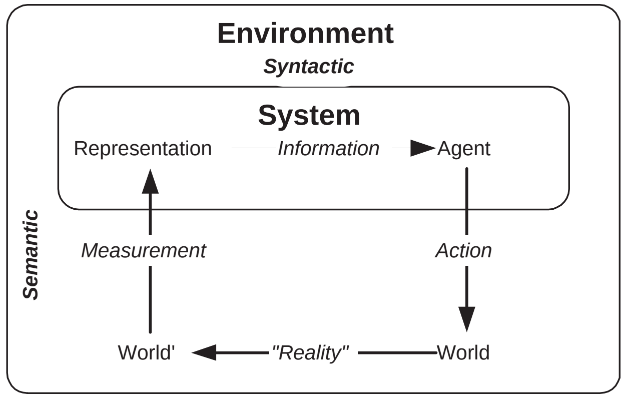

Figure 1: A semiotic control system.

of the environment by the system is mediated through the measurement (perception) process, which provides a partial representation of the environment to the system. In turn, the system takes certain actions into the world, but still has no knowledge of the consequences of those actions except as reflected in later measurements.

Between the system's measured state and the agent which decides (performs) its action is what can be called an information or knowledge relation. Minimally, this is just the transmission of the representation of the environment to the agent. In more complex systems, this plays the role of cognition, information processing, or knowledge development. Typically, extra or external knowledge about the state of the world or the desired state of affairs is brought to bear, and provided to the agent in some processed form, for example as an error condition or distance from optimal state.

Clearly control systems are involved in cyclic relations with their environments. To be in good control, the overall system must form a negative feedback loop, so that disturbances and other external forces from "reality" (for example noise or the actions of other external control systems) are counteracted by compensating actions so as to make the measured state as close as possible to some desired state. If rather a positive feedback relation holds, then such fluctuations will be amplified, ultimately bringing some critical internal parameters beyond tolerable limits, or otherwise exhausting some critical system resource, and thus leading to the destruction of the system as a viable entity.

2.1 Control as a Semantic Relation

We interpret the control system in a semiotic context, thereby evoking the distinctions among syntactic (the formal properties of symbol tokens as used in symbol systems), semantic (the interpretation of tokens as their meanings), and pragmatic (the use of symbol tokens and their meanings for the overall purposes or survivability of the system) relations [1, 3].

The processes in the control relation, and especially the decision-making functions of the agent, are inherently semantic. The role of the agent is to accept information about the measured state of the environment, compare it to the desired state, and, based on the discrepancy, to engage in a decision-making process of selecting one action amongst multiple possible actions which is appropriate for the goal of maintaining the overall system in a stable, negative feedback relation.

This making of appropriate choices is exactly the semantic function in a semiotic system. It is on this required "appropriateness" of the choice of the agent that the "intelligence" of the semiotic system rests: a certain action is "correct" in a given context, while another is not. This semantic function is a coding relation, which is a contingent entailment, meaning that it is arbitrary, but fixed [10].

This is the hallmark property of semantic systems, that the coding of their symbol tokens act as contingent functional entailments, and are thus dually contingent and necessary at complementary levels of analysis. That is, from *within* the symbol system, the token must necessarily be interpreted according to the code, but from *without* we are (or "evolution" is) free to choose any coding we please. They are conventional, constructed and interpretable by a certain closed "linguistic community" [16].

Thus codes act as *rules* within systems (to use Pattee's language [18]), as contrasted with *natural laws*, which are wholly necessary at all levels of analysis. This combination of freedom and determinism is not possible with purely physical systems. Indeed, the

X	$f_1(X)$	$f_2(X)$	$f_3(X)$
$+$	d	d	d
$-$	u	u	u
0	n	d	u

Table 1: Functions sufficient for semiotic control.

school of biosemiotics [4] is dedicated, in some sense, to the proposition that the classes of semiotic systems and living systems are equivalent, or at least coextensive [4, 10].

A simplified example will serve to illustrate this point. Let O be a simple organism which lives near an oceanic thermocline with warm water above and cold water below. O acts as a semiotic control system in relation to the thermocline. Its perception is a single critical variable of temperature with states

$X = \{+ = \text{too hot}, - = \text{too cold}, 0 = \text{just right}\}$,

and it has a single variable action with states

$Y = \{u = \text{go up}, d = \text{go down}, n = \text{do nothing}\}$.

The information relation is simply transmission of X to the agent.

There are $3^3 = 27$ possible functions $f : X \mapsto Y$, any of which the agent could invoke to make a decision to take a particular action, but only the three shown in Tab. 1 will result in stable negative feedback control. f_1 is the best default selection, since it minimizes unnecessary action and results in smoother and faster control. But if f is not selected from these three, then positive feedback, not negative feedback, will result, with a corresponding runaway behavior.

There is no fundamental natural law of the universe which requires f to be selected according to the principles of negative feedback. Instead, this selection is *contingent on*, and *results from*, the process by which the system is *constructed*.

In a system which has contingent entailments (rules), there are semantic relations of meaning among the components. In our example, it is appropriate to say that for our organism "too hot" actually *means* "go down", and "too cold" actually means "go up".

2.2 Decision-Making as an Informational Process

Not only are semantic relations present in semiotic systems, but these can be measured in information theoretical terms. Consider a (finite) set of possibilities $\Omega = \{\omega\}$ available to the agent. Classically, the process of selection can be described as the identification of a specific choice $\omega^* \in \Omega$.

Consider the agent as a communication channel. Before the choice is made, there are $|\Omega|$ possibilities, and thus there is an inherent *a priori* uncertainty as to which choice $\omega^* \in \Omega$ will be selected. Furthermore, after the selection is made, there is total certainty as to what the choice was, namely ω^*. Therefore we can regard the process of decision-making performed by the agent as the reduction of uncertainty as to which choice *will be* made, and conversely as the gain of information as to which choice *was* made. In either case the amount of this uncertainty or information is quantified as $I = \log_2(|\Omega|)$.

In the classical case the set of available choices is constrained to exactly one: $\omega^* \in \Omega$. To express the selection of the action as a weaker form of constraint, we can instead identify a subset $\emptyset \neq \Omega^* \subseteq \Omega$, representing now not a particular choice, but a nondeterministic restriction of the possible choices. Now the information is quantified as $\log_2(|\Omega|/|\Omega^*|) = \log_2(\Omega) - \log_2(\Omega^*)$, where the prior situation remains a special case where $\Omega^* = \{\omega^*\}$.

3 General Information Theory Methods

All of this is well known, and has been extended in a variety of ways in information theory. Classical information theory is based on situations where the choices the agent makes are quantified according to the rules of probability theory, and entropy measures are used to quantify the total quantity of uncertainty or information involved.

But recent years have seen a proliferation of mathematical methods for the quantification of these choices, and a similar proliferation of global measures of information. Most of these methods attempt to generalize away from probability in one way or another, but remain intimately related to it.

3.1 Generalized Distributions and Measures

One important generalization is to define decision-making as the identification of a fuzzy subset $\tilde{A} \tilde{\subseteq} \Omega$, where $\tilde{A}$ is defined by its membership function $\mu_{\tilde{A}} : \Omega \mapsto [0, 1]$. Nondeterministic constraint remains a special case where $\forall \omega \in \Omega, \mu_{\tilde{A}}(\omega) \in \{0, 1\}$.

For any possibility $\omega \in \Omega$, the value $\mu_{\tilde{A}}(\omega)$ indicates the weight placed on the particular choice ω. If $\mu_{\tilde{A}}(\omega) = 0$ then ω is eliminated; if $\mu_{\tilde{A}}(\omega) = 1$ then ω is completely allowed; and if $\mu_{\tilde{A}}(\omega) \in (0, 1)$ then ω is partially allowed. The classical case is expressed by $\exists! \omega^* \in \Omega, \mu_{\tilde{A}}(\omega *) = 1, \forall \omega \neq \omega^*, \mu_{\tilde{A}}(\omega) = 0$.

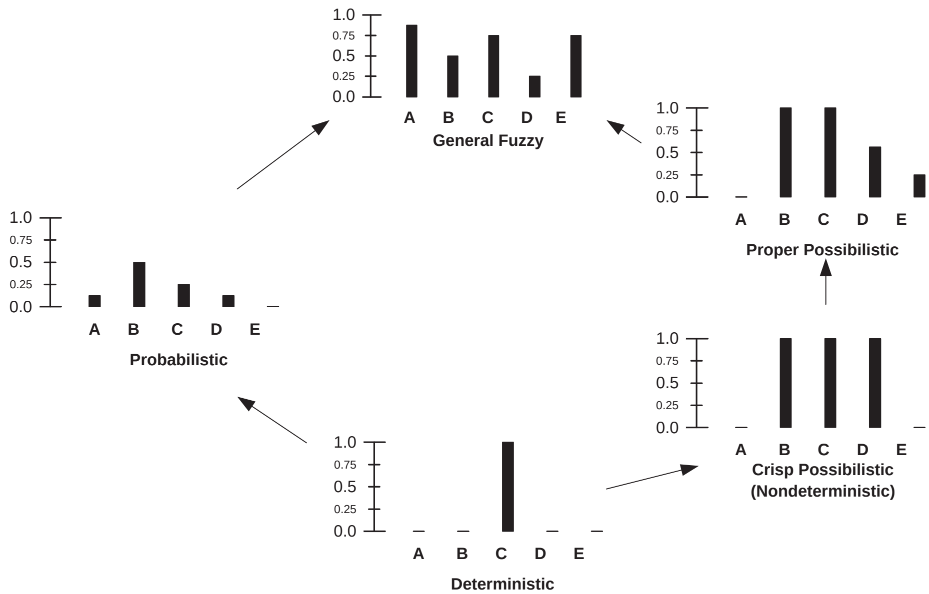

Figure 2: Types of distributions.

So taken as a whole, the entire function $\mu_{\tilde{A}}$ assigns a set of arbitrary weights to choices. It may be that $\mu_{\tilde{A}}$ has some other interesting mathematical properties so that it acts as a kind of generalized distribution. In particular, we are interested in properties of the distribution with respect to operators of the form $\oplus\colon [0,1]^2 \mapsto [0,1]$, usually taken as addition-like triangular conorms $\sqcup\colon [0,1]^2 \mapsto [0,1]$ (commutative, associative, and monotonic, with identity 0). Given a conorm $\sqcup$, then a fuzzy set $\mu_{\tilde{A}}$ is said to be $\sqcup$-normal if $\bigsqcup_{\omega \in \Omega} \mu_{\tilde{A}}(\omega) = 1$.

There are many parameterized families of conorm operators, but the ones of most interest to us here are the bounded sum $a +_b b := (a+b) \wedge 1$, where $a, b \in [0,1]$ and $\wedge$ is minimization, and the the maximum operator $a \vee b$. Since normalization under the operator $+$ (not itself a conorm, since $+\colon [0,1] \mapsto [0,2]$) implies normalization under $+_b$, therefore a probability distribution p is a special $+_b$-normal fuzzy set. A $\vee$-normal fuzzy set is called possibility distribution π. The measure of information then becomes the classical entropy measure for probability theory, and the nonspecificity measure for possibility theory

$$\mathbf{H}(p) \quad := \quad -\sum_{\omega \in \Omega} p(\omega) \log_2 p(\omega),$$

$$\mathbf{N}(\pi) \quad := \quad \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i)$$

respectively, where in the possibilistic case the π_i are ordered with $1 \leq i \leq n$ and $\pi_{n+1} := 0$.

Possibility theory [2] provides an important new form of information theory, and is crucial because it generalizes nondeterminism and nonspecificity in a far superior way to probability theory. Nondeterministic constraint is identified as a case of crisp possibility, where $\forall \omega \in \Omega, \pi(\omega) \in \{0,1\}$, so that $\Omega^* = \{\omega : \pi(\omega) = 1\}$. Similarly, $\mathbf{N}(\pi)$ generalizes the $\log_2(|\Omega^*|)$ measure in this case. In this way, for example, fuzzy intervals (convex, normal fuzzy sets of $\mathbb{R}$) properly generalize classical intervals, and possibilistic processes properly generalize state transitions with an intermediate degree of nondeterminism [8, 12].

As fuzzy sets generalize probability distributions, so fuzzy measures generalize probability measures. Given a $\sqcup$-normal fuzzy set $\mu_{\tilde{A}}$, then a normal $\sqcup$-distributional ($\sqcup$-decomposable) finite fuzzy measure (a monotonic set function $\nu\colon 2^\Omega \mapsto [0,1]$ [21]) results from the construction

$$\nu(A) := \bigsqcup_{\omega_i \in A} \mu_{\tilde{A}}(\omega), \qquad A \subseteq \Omega.$$

When $\mu_{\tilde{A}}$ is $+$-normal, then ν is a probability measure with the usual additive properties, and when $\sqcup = \vee$ then $\Pi := \nu$ is a possibility measure with the maximal property

$$\Pi(A \cup B) = \Pi(A) \vee \Pi(B), \qquad A, B \subseteq \Omega.$$

Similarly, the distributions (fuzzy sets) are recovered from measures by the relation $\mu_{\tilde{A}}(\omega) = \nu(\{\omega\})$.

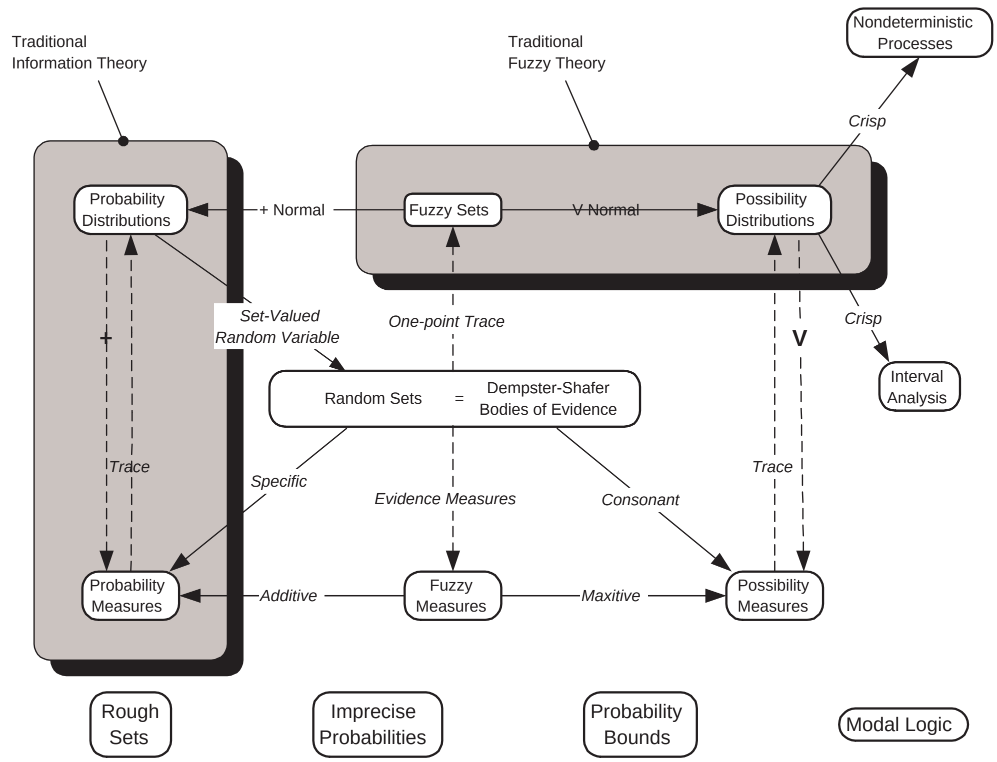

Figure 3: (Top) Relations among some of the components of GIT. (Bottom) Other components being investigated.

3.2 Random Sets and Dempster-Shafer Evidence Theory

An important goal for the GIT research community is to synthesize these methods in order to find common axiomatizations, and random sets provide one of the most satisfying mathematical forms to do so. Random sets generalize random variables in that they are set-valued, and are simple hybrid structures combining probabilistic randomness with the nonspecificity of variable subset size, extent, and overlap.

Given a focal set (a collection of subsets $\mathcal{F} = \{A_j : \emptyset \neq A_j \subseteq \Omega\}, 1 \leq j \leq N$), each is assigned a probability of occurrence $m(A_j) \in [0,1]$ with $\sum_{A_j \in \mathcal{F}} m(A_j) = 1$. Then $\mathcal{S} := \{\langle A_j, m(A_j)\rangle\}$ is a random set [13].

These statistical structures are mathematically isomorphic to Dempster-Shafer bodies of evidence. They do not yield to traditional statistical interpretations in that a unique probability distribution p on Ω consistent with the given evidence m cannot generally be recovered. Rather, they generate the (usually non-distributional) fuzzy measures Bel and Pl given by

$$\text{Bel}(A) := \sum_{A_j \subseteq A} m(A_j), \qquad \text{Pl}(A) := \sum_{A_j \cap A \neq \emptyset} m(A_j),$$

which act as bounds on the probability measure associated with the hypothetical distribution p, and a fuzzy set called the one-point trace with membership $\rho(\omega) := \text{Pl}(\{\omega\})$.

There are two important special cases of the topology of the focal set: when the A_j are all singletons, then $\mathcal{S}$ is specific, and Pl = Bel is a probability measure with distribution $p := \rho$; and when the A_j are all nested with $A_{j-1} \subseteq A_j$ ($A_0 := \emptyset$), then Pl is a possibility measure with distribution $\pi := \rho$. A map of all of these structures and their relations is shown in the top of Fig. 3.

3.3 Other Components of GIT

There are a number of other mathematical formalisms which have been demonstrated to be closely related to the above cluster, or whose connections are being actively pursued.

Rough sets [19] are structures which approximate a given statement by an inner and an outer approximation, and in their simple forms are also crisp possibility distributions. Walley's imprecise probabilities [20] generalize probability measures directly, and generate a variety of interesting structures including evidence measures. Probability bounds on cumulative distribution functions [5] are in some ways similar to evidence measures, in that they operate on families of distributions, but their connection to fuzzy measures remains unclear. Finally, there has been recent work to interpret modal logic in the context of evidence measures, including possibility theory [7].

4 Consequences

So where does the recognition of this wide array of information theoretical methods leave us with respect to decision making in the context of semiotic systems?

4.1 Formalisms and Applications

The first consideration is that from the semiotic perspective, just as the relation of symbols to their referents is a contingent entailment, so also is the relation of mathematical theories to their applications and interpretations. The formal, syntactic properties of mathematical systems follow from their axioms, but we are free to interpret them in any particular context we choose: *mathematical theories do not come with inherent semantics.*

This has been a problem, for example, in fuzzy systems theory, where it has long been presumed that fuzzy sets *necessarily* represent human cognitive assessments of uncertainty through linguistic categories. No doubt that is an important application, but only one among many.

In general we have many mathematical theories with mappings and transformations among them and many semantic domains of application, some of which have representations in some of those theories. When a particular method is used for a particular application, and that method has mappings to a different method, then it is suggested, if not demanded, that the original interpretation be reconsidered from the perspective of the new formalism (see Fig. 4).

Such is the case, for example, among random sets, evidence theory, and fuzzy measure theory. While random set and evidence theory are formally equivalent, they have very different histories and agendas; evidence theory is primarily a cognitive modeling method, while random set theory has been used for statistical geometry, data fusion, and image analysis. Similarly, fuzzy measures are widely used in knowledge engineering applications, but their corresponding random set interpretations are not generally pursued.

4.2 Measurement and Action Methods

The effects of this are seen where the semantic functions are involved. That is, our choice of interpretations will be dependent on the available measurement and action methods. For example, in classical information theory, histograms and sample statistics are used to generate measured probability distributions by statistical inference. And in classical fuzzy systems, non-statistical knowledge elicitation methods are used to measure the cognitive state of a human subject.

So non-subjective interpretations of fuzzy sets and fuzzy measures must wait on the development of non-subjective measurement procedures. Such procedures are provided by random set measurement methods [11], which promise, for example, to provide the best *empirical* basis for possibility theory.

Similarly, appropriate measurement and action methods are necessary for possibility theory, rough set theory, etc. Such movement is under way, and following our previous argument, direction can be taken by drawing on formal analogies from other mathematical theories. For example, in this way possibilistic Monte Carlo methods [9] and suggestions about potential sample statistics [22] have been advanced.

Ultimately, the development of scientific theories is dependent on the formal languages available to them: new formal systems allow the possibility of new forms of measurement from and action into the world. Similar processes have occured in organisms over biological time, in the form of the evolution of new modes of perception and action. It can only be hoped that further mathematical development will ultimately open new vistas for scientific and engineering applications as well.

References

[1] Cariani, Peter A: (1989) *On the Design of Devices with Emergent Semantic Functions*, SUNY-Binghamton, Binghamton NY, PhD Dissertation

[2] de Cooman, Gert; Ruan, D; and Kerre, EE, eds.: (1995) *Foundations and Applications of Possibility Theory*, World Scientific, Singapore

[3] Deely, John: (1990) *Basics of Semiotics*, Indiana UP, Bloomington IN

[4] Deely, John: (1992) "Semiotics and Biosemiotics: Are Sign-Science and Life-Science Coextensive?", in: *Biosemiotics: The Semiotic Web 1991*, ed. TA Sebeok, J Umiker-Sebeok, pp. 46-75, Mouton de Gruyter, Berlin/NY

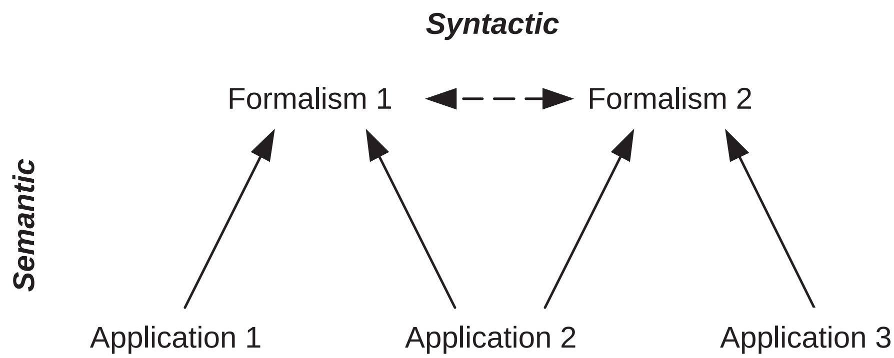

Figure 4: Relations among formalisms and their applications.

[5] Ferson, Scott; Ginzburg, Lev; and Akcakaya, Resit: (1996) "Whereof One Cannot Speak: When Input Distributions are Unknown", *Risk Analysis*, in press

[6] Guan, J and Bell, DA: (1992) *Evidence Theory and its Applications*, v. **1-2**, North Holland

[7] Harmanec, David; Klir, George; and Resconi, Germano: (1995) "On Modal Logic Interpretation of Dempster-Shafer Theory of Evidence", *Int. J. General Systems*

[8] Janssen, Hugo J; de Cooman, Gert; and Kerre, Etienne: (1996) "First Results for a Mathematical Theory of Possibilistic Processes", in: *Proc. 1996 European Meeting on Cybernetics and Systems Research*, pp. 341-346, ed. Robert Trappl

[9] Joslyn, Cliff: (1994) *Possibilistic Processes for Complex Systems Modeling*, PhD dissertation, SUNY-Binghamton, UMI Disseration Services, Ann Arbor MI

[10] Joslyn, Cliff: (1996) "Are Meaning and Life Coextensive?", in: *Proc. 1995 Int. Symposium on Evolving Systems*, in press

[11] Joslyn, Cliff: (1996) "Measurement of Possibilistic Histograms from Interval Data", *Int. J. General Systems*, in press

[12] Joslyn, Cliff: (1996) "The Process Theoretical Approach to Qualitative DEVS", in: *Proc. 7th Conf. on AI, Simulation, and Planning in High Autonomy Systems (AIS '96)*, ed. Zeigler, Bernie, pp. 235-242, San Diego CA

[13] 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

[14] Klir, George: (1991) "Generalized Information Theory", *Fuzzy Sets and Systems*, v. **40**, pp. 127-142

[15] Klir, George and Yuan, Bo: (1995) *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall, New York

[16] Lewis, David: (1969) *Convention*, Harvard U Press, Cambridge

[17] Meystel, Alex: (1995) *Semiotic Modeling and Situation Analysis: An Introduction*, AdRem Inc., Bala Cynwyd PA

[18] Pattee, Howard: (1982) "Cell Psychology: An Evolutionary View of the Symbol-Matter Problem", *Cognition and Brain Theory*, v. **5**, pp. 325-341

[19] Pawlak, Zdzislaw: (1991) *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Boston

[20] Walley, P: (1990) *Statstical Reasoning with Imprecise Probabilities*, Chapman and Hall, New York

[21] Wang, Zhenyuan and Klir, George J: (1992) *Fuzzy Measure Theory*, Plenum Press, New York

[22] Wierman, Mark J: (1996) "Assessing Fuzzy Sets and the Potential of Possibility Theory", in: *Information Sciences*, v. **88**, pp. 247-261