A Qualitative Framework for Shannon Information Theories

Dr. Gerard Allwein
Center for High Assurance Computer Systems, Code 5540
Naval Research Laboratory
Washington, D.C. 20375 USA
allwein@itd.nrl.navy.mil

ABSTRACT

This paper presents a new paradigm for information theory which is a synthesis of Barwise-Seligman's qualitative theory and Shannon's quantitative theory. The new paradigm is best viewed as a meta-theory for Shannon information theories and allows different probability theories, and subsequently, new Shannon information theories, to work within a common framework. The resulting Shannon theories conform to a qualitative structure and decorate it with measures of information. This approach is useful for analyzing assurance problems where there the analysis must contend with incomplete and even contradictory information. In particular, the mathematical constructs of the theory allow one to use just about any logic which admits a companion measure theory.

Keywords

Barwise-Seligman Information Theory, Shannon Information Theory

1. INTRODUCTION

Information is widely viewed as something in need of control. However, information viewed as "data" or "what moves on a wire" is much too crude a notion to be useful in serious information assurance matters. Barwise and Seligman in [5] and Dretske [6] propose that information is not reducible to something as self-explanatory as data or the substance of communication. The situation is best viewed with an example. Suppose there is some sensitive information hidden on some computer. What does it mean to say that the information has leaked? If the answer is merely that some attack was successful in transferring the bits representing the information to an outside agency, then presumably the bits contain all there is to the information. This cannot be the case as Ruth Nelson observed in [11].

The obvious first reply is, the information is encoded, hence whomever has the data does not necessarily have the information. Suppose the key to the encoding is also leaked. Does the recipient now have the information? Let the information be a (now un-encoded) single bit representing the answer to a yes or no question. Receiving the value of the bit does not entail one knows the information. What is missing is the context. Including the context in the information along with the bit only begs the question of what context is needed in order to understand the context and an infinite regress ensues.

The answer to "What is information?" greatly influences what methods one employs in either protecting or diagnosing an attack on information. Barwise-Seligman's definition of information is anything of the form "x is an A" where x can be thought of as a piece of data and A is something that can be said about that piece of data. That said, the deeper more important question is (from [5]) "How *do* remote objects, situations, and events carry information about one another without any substance moving between them?" They propose a mathematical framework for answering this question. Their book's subtitle, "The Logic of Distributed Systems" hints at the features for which their framework is built: (1) information comes in the form of distributed systems within which remote parts carry information about other parts, and (2) it is logic that ties the distributed system together. Upon this framework, communication, assurance, and a host of other properties can be modeled.

The new paradigm presented in this paper can also be seen as direct response to Ruth Nelson's quest for better models. As she also observed, confidential information is not something admitting a generic abstract definition. However, one can use the new theory presented here as an architecture for modeling situations in which confidential information is present, how it is to be protected, and how to diagnose that it is indeed protected.

Shannon in [14], "Communication Theory of Secrecy Systems" makes note of (1) concealment systems, (2) privacy systems, and (3) cryptographic systems. He dismisses the first two kinds as psychological and technological respectively. And we, respectfully, and vehemently disagree. It is because of this kind of mindset from communication specialists that "secrecy" as a non-functional property **capable of formal presentation** has only recently been taken seriously. Recourse to defining a secret as something to which only zero-capacity channels are connected will not work, one can send an arbitrary amount of information through a zero-capacity channel, see [9].

In [9], a new paradigm is presented for steganography. They raise the issue that merely measuring the content of a hidden image is not sufficient if nothing is known about

NSPW 2004 Nova Scotia Canada

the typing scheme used in the encoding. Barwise-Seligman theory answers this directly in their definition of information since no piece of data becomes information without a typing scheme. Forgetting steganography and considering the mere sending of a map through a channel, when do enough of the map's relationships make it through the channel so that information can be claimed to be received? It is related to the noise in the channel, but not directly. If two buildings are fuzzy in the received map but can still be discerned as being across from each other, that might be enough to claim the information has been received. The point is that there is a qualitative nature to information that cannot be ignored.

Barwise-Seligman's qualitative framework is quite general. However, it suffers from no provisions for the measurement of information. Consequently, it is sometimes difficult to say when a particular information flow is *probable*. It is vital to assign measurements so that resources can be directed to where the biggest payoffs lie.

Claude Shannon [15] proposed a communication theory that Dretske and many before him have attempted to coerce into a theory of information. Their efforts generally fail because while one can put measures on data flow, data flow is not necessarily information flow. Dretske's work is easily the most sophisticated of these efforts and he does come close to treating information in its more subtle senses, but it does not include a wide enough mathematical framework that includes what is the most paradigm use of information, that we *reason* with it. Channel theory is such a framework.

Our new paradigm combines Shannon's and subsequently, Dretske's, quantitative theory with Barwise and Seligman's qualitative theory. We believe it provides a good theoretical foundation in which to view information assurance problems. In particular, our new paradigm allows:

- The use of specialized logics where they make sense.
- The use of measures spawned by specialized logics so that information flow, either as communication or as mere flow of reasoning, can be measured.
- A rigorous, composable mathematical structure that includes provisions for combining fine grained analyses into larger grained analyses, i.e., it scales well.

Due to paper length, the logic-plus-measure theory used here will be classical propositional logic and classical probability theory. The method used to develop the theory here is similar for any logic-plus-measure theory as our current research shows [2] [3].

2. SAMPLE PROBLEMS

These problems will be revisited at the end of the paper to show what a solution would look like in our new paradigm.

Capacity Issues Is it possible to take a simple Shannon-type problem and derive the Shannon-type solutions using logically derived methods? Can the solution be put in a form that uses the same theory as the other examples of this section?

BLP A low priority clearance should not allow reading of high priority information. And high priority information should not be written down to low priority information. Can this be modeled with logical rules and then measures placed on the information flows within the system?

Quasi-Coordinated Attack Let $\boldsymbol{C}$ refer to a computer system which is attacked from several different directions represented by $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$, some of them coordinated. Let $\boldsymbol{B}$ be a system resource which is attacked. The problem is to define possible scenarios, or better theories, which show how the $\boldsymbol{A}_i$'s are connected to each other and to $\boldsymbol{B}$. Measures must be placed on the theories showing which are most probable. It must be possible to reason forward from the $\boldsymbol{A}_i$'s through $\boldsymbol{C}$ to $\boldsymbol{B}$ and backward from $\boldsymbol{B}$ to the $\boldsymbol{A}_i$'s through $\boldsymbol{C}$ (this latter is called "reasoning by abduction").

3. CHANNEL THEORY

The generic term used to describe Barwise-Seligman's information theory is *channel theory* and is called by them "the logic of distributed systems" for a very good reason. It shows how local logics in the distributed parts of a system are connected via logics in informational connections or channels. The channels are not necessarily communication channels (but they can be if the type of problem you wish to solve requires that they be), they are much more general than that. Our research expands channel theory by providing probability theories associated to these logics and hence make the result amenable to suitable Shannon-type information theories. The expanded theory will also be called *channel theory* and it is this expanded theory that is used in the sequel.

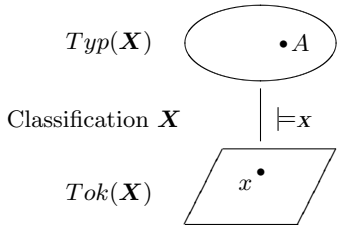
It is also of worth to note that Shannon has a little recognized and fairly abstruse paper [13] where he tiptoes up the notion of a channel (of channel theory) but never really takes the plunge.

3.1 Basic Structures

The basic structures of channel theory are deceptively simple. The things that are distributed in a distributive system are *contexts* called *classifications*. The classifications are connected by *infomorphisms*. The relevant definitions follow:

Definition 3.1.1 (Barwise–Seligman) *A* **classification**, $\boldsymbol{X}$, *is a pair of sets and a relation. The sets are called, respectively, the* **tokens**, $Tok(\boldsymbol{X})$, *and* **types**, $Typ(\boldsymbol{X})$. *The binary relation, usually symbolized by* $\models_{\boldsymbol{X}}$, *is between the two sets, i.e.,* $\models_{\boldsymbol{X}} \subseteq Tok(\boldsymbol{X}) \times Typ(\boldsymbol{X})$. *The term* $x \models_{\boldsymbol{X}} A$ *means* $\langle x, A\rangle \in \models_{\boldsymbol{X}}$ *with* $x \in Tok(\boldsymbol{X})$ *and* $A \in Typ(\boldsymbol{X})$.

A good mental picture to remember the definition is the following:



The diagram only indicates that $x \in Tok(\boldsymbol{X})$ and $x \in Typ(\boldsymbol{X})$, not that $x \models_{\boldsymbol{X}} A$.

It is convenient to talk about all of the tokens satisfying a single type or all of the types satisfying a particular token. The following definition relativizes $Typ(-)$ and $Tok(-)$ to a particular classification.

Definition 3.1.2 *Let* $\boldsymbol{X} = (Tok(\boldsymbol{X}), Typ(\boldsymbol{X}), \models_{\boldsymbol{X}})$ *be a classification, then for any* $A \in Typ(\boldsymbol{X})$, $Tok(A) = \{y \mid y \models_{\boldsymbol{X}} A\}$ *and, for any* $x \in Tok(\boldsymbol{X})$, $Typ(x) = \{B \mid x \models_{\boldsymbol{X}} B\}$.

It is frequently helpful to define the following preorder on types of a classification:

Definition 3.1.3 *Given a classification* $\boldsymbol{X}$, *the* **token induced preorder** *on* $Typ(\boldsymbol{X})$ *is defined with*

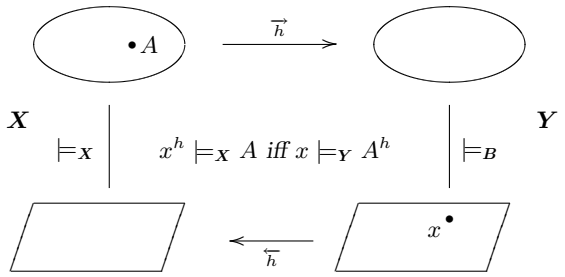
$$A \prec B \text{ iff } Tok(A) \subseteq Tok(B).$$

The reason $\prec$ is a preorder instead of a partial order is because the collection of types do not necessarily have an extensional character, i.e., they need not be sets. If your types are formulas and tokens are interpretations, two formulas A and B can have the same interpretations which make them simultaneously true and false, hence their tokens sets are equal. But as sentences, A and B can be very different. An analogous preorder on tokens exists. The infomorphisms defined next preserve these preorders.

Definition 3.1.4 (Barwise–Seligman) *Assume that* $\boldsymbol{X} = (Tok(\boldsymbol{X}), Typ(\boldsymbol{X}), \models_{\boldsymbol{X}})$ *and* $\boldsymbol{Y} = (Tok(\boldsymbol{Y}), Typ(\boldsymbol{Y}), \models_{\boldsymbol{Y}})$ *are classifications. An* **infomorphism** $h : \boldsymbol{X} \longrightarrow \boldsymbol{Y}$ *is a pair of contravariant maps,* $\overrightarrow{h}$ *and* $\overleftarrow{h}$ *such that* $\overrightarrow{h} : Typ(\boldsymbol{X}) \longrightarrow Typ(\boldsymbol{Y})$ *and* $\overleftarrow{h} : Tok(\boldsymbol{Y}) \longrightarrow Tok(\boldsymbol{X})$, *and for all* x *and* A, *the following condition is satisfied:*

$$x^h \models_{\boldsymbol{X}} A \text{ iff } x \models_{\boldsymbol{Y}} A^h,$$

where for ease of presentation, $\overleftarrow{h}(x)$ *is displayed as* x^h *and* $\overrightarrow{h}(A)$ *as* A^h. *This can be pictured with:*



3.2 Classical Propositional Logic: Example

The following two definitions single out Countable Classical Propositional Logic (CCPL). CCPL is classical proposition logic outfitted with countable conjunction and disjunctions such that the algebraic models for the logic are σ-algebras. It turns out the topological spaces dual the algebras have the same points as the Stone space duals to the underlying Boolean algebras (of the σ-algebras).

Definition 3.2.1 *A* ***Stone interpretation*** *for CCPL is any interpretation* x *which satisfies the following conditions on a relation* $\models$ *for* A *either a formula or event and* Γ *either a set of formulas in the language of CCPL or a set of events in a* σ*-algebra:*

I1: $x \models \neg A$ iff $x \not\models A$;

I2: $x \models \bigwedge \Gamma$ iff $x \models A$ *for all* $A \in \Gamma$;

I3: $x \models \bigvee \Gamma$ iff $x \models A$ *for some* $A \in \Gamma$.

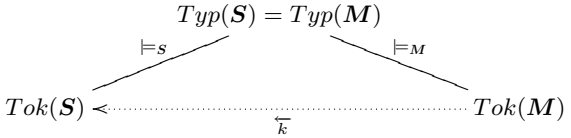
Definition 3.2.2 *The category of CCPL is a subcategory of all classifications and infomorphisms such that:*

- $Typ(\boldsymbol{M}) \stackrel{def}{=}$ *formulas of CCPL or events of a* σ*-algebra;*
- $Tok(\boldsymbol{M})$ *is a collection of objects satisfying conditions I1-I3 of being a Stone interpretation and* $Tok(\boldsymbol{S}) \neq \emptyset$;
- $\models_{\boldsymbol{M}}$ *any relation that properly relates Stone interpretations to CCPL formulae according to I1-I3.*

In the classifications for CCPL, $\prec$ becomes the logic's entailment order. If the types were elements of a σ-algebra, $\prec$ becomes the σ-algebra's lattice order. The following theorem allows the use of token sets, which are sets of occurrences, to be used.

Theorem 3.2.3 *There is a free object,* $\boldsymbol{S}$, *in the category of CCPL classifications.*

The freeness of the situation is as follows where $Typ(\boldsymbol{S})$ is either formulas of CCPL or a σ-algebra of events. The tokens of $\boldsymbol{S}$ are the canonical Stone interpretations. The tokens of any other object $\boldsymbol{M}$ are occurrences which act just like Stone interpretations except that repeated elements are allowed. It is these repeated elements that cause types in $\boldsymbol{S}$ to achieve a weight measured by a probability function. The infomorphism k is the identity function on types and maps tokens into the "states" of the canonical collection of interpretations. $\overleftarrow{k}$ appears as a co-free morphism if only the token sets are considered. By convention, the infomorphism k goes in the opposition direction as the token map, $\overleftarrow{k}$.



A probability function, P, is intimately connected with the classification $\boldsymbol{M}$. The unique morphism k is able to map some tokens into the same Stone interpretation. If the number of interpretations were finite, i.e., the language of CCPL or set of events were finite, one could assign a weight to each A based on the characteristics of k. These weights would then be combined to yield P. However, when the number of interpretations is infinite, then P is related to k but cannot be derived from k by summation. Instead, a frequency analysis must be used. The point remains that k and P are intimately related. P should really have a subscript, i.e., P_k.

4. SEQUENTS AND LOGICS

A *sequent* represents a constraint that may or may not hold of a classification. It is a logical statement in that it represents a relation between premises and conclusions. The premises and conclusion are sets of types. It is sequents that enable the flow of information. The information flow they enable is an information flow of reasoning. That said, sequents may also be used to model communication flows when the sequents are modeling communication. A communication sequent or *gate* can be thought of as allowing a token to flow under it just when the token satisfying the premises also entails that the token satisfy the conclusion.

4.1 Sequents

Definition 4.1.1 (Barwise–Seligman) *Let $\boldsymbol{A}$ be a classification. A* **theory** *for $\boldsymbol{A}$ is a collection of* **sequents** *of the form:*

$$\Gamma \Vdash_{\boldsymbol{A}} \Delta$$

where Γ and Δ are collections of types and the $\Vdash_{\boldsymbol{A}}$ is the turnstile of logical consequence.

This is the usual notion of sequent. The types in Γ are thought of as conjoined together and the types in Δ are thought of as disjoined. The requirement for a token, x, to satisfy the above sequent is:

$$\begin{array}{c}(\text{for all } P \in \Gamma, x \models_{\boldsymbol{A}} P) \text{ implies} \\ (\text{there exists one } Q \in \Delta, x \models_{\boldsymbol{A}} Q).\end{array}$$

When Γ or Δ are singleton sets, say, $\{A\}$, then $A \Vdash_{\boldsymbol{A}} \Delta$ or $\Gamma \Vdash_{\boldsymbol{A}} A$ will be used. It is important to notice there is no logical structure imposed on the types as a restriction imposed by channel theory. They are merely types. Any logical structure could be imposed as a result of attempting to model some domain of discourse, but channel theory *simpliciter* does not impose one itself. Any extra structure would come about because some peculiar feature of a universe of discourse needed to be modeled.

4.2 Logics

Definition 4.2.1 (Barwise-Seligman) *A* **local logic** $\mathcal{L} = \langle \boldsymbol{A}, \Vdash_{\mathcal{L}}, N_{\mathcal{L}} \rangle$ *consists of a classification $\boldsymbol{A}$, a set $\Vdash_{\mathcal{L}}$ of sequents involving the types of $\boldsymbol{A}$, and a subset $N_{\mathcal{L}} \subseteq Tok(\boldsymbol{A})$ called the* **normal tokens** *of $\mathcal{L}$, which satisfy all the constraints $\Vdash_{\mathcal{L}}$. A local logic $\mathcal{L}$ is sound if every token is normal; it is complete if every sequent that holds of all normal tokens is in the consequence relation $\Vdash_{\mathcal{L}}$.*

Typically, the sequents are required to follow certain *structural rules* but these will not concern us in this paper. The following two (non-structural) rules allow for the movement of logics between classifications connected via the infomorphism $f : \boldsymbol{A} \longrightarrow \boldsymbol{B}$:

$$\frac{\Gamma^{-f} \Vdash_{\boldsymbol{A}} \Delta^{-f}}{\Gamma \Vdash_{\boldsymbol{B}} \Delta}\ f\text{-Intro} \qquad \frac{\Gamma \Vdash_{\boldsymbol{A}} \Delta}{\Gamma^{f} \Vdash_{\boldsymbol{B}} \Delta^{f}}\ f\text{-Intro}$$

$$\frac{\Gamma^{f} \Vdash_{\boldsymbol{B}} \Delta^{f}}{\Gamma \Vdash_{\boldsymbol{A}} \Delta}\ f\text{-Elim} \qquad \frac{\Gamma \Vdash_{\boldsymbol{B}} \Delta}{\Gamma^{-f} \Vdash_{\boldsymbol{A}} \Delta^{-f}}\ f\text{-Elim}$$

where Γ^{-f} is a nicer way of writing $\overrightarrow{f}^{-1}(\Gamma)$, i.e., the inverse image of Γ under f and Δ^{f} is the direct image of Δ under f. Each rule has two forms. f-Intro preserves validity, to wit: assume the premise and let x be a counter-example to the conclusion. If $x^{f} \models_{\boldsymbol{A}} P$ for all $P \in \Gamma^{-f}$ (vacuously if $\Gamma^{-f} = \emptyset$), then x^{f} must satisfy at least one $Q \in \Delta^{-f}$. Since $x^{f} \models_{\boldsymbol{A}} Q$, then $x \models_{\boldsymbol{B}} Q$ which is a contradition to x being a counter-example. f-Elim fails to observe validity since it is possible for a counter-example in the conclusion to have no preimage under $\overleftarrow{f}$. Of course, if $\overleftarrow{f}(Tok(\boldsymbol{B})) = Tok(\boldsymbol{A})$, then the rule will preserve validity. Preservation of non-validity is exactly the opposite for the two rules.

The two different forms of the rules are quite different because they are working on sets. Consider the two cases of f-Elim. In the first, the types in Γ and Δ are types of $\boldsymbol{A}$ that have been mapped to $\boldsymbol{B}$ under f. In the second, the types in Γ and Δ are types of $\boldsymbol{B}$ that are pulled back along f to types in $\boldsymbol{A}$.

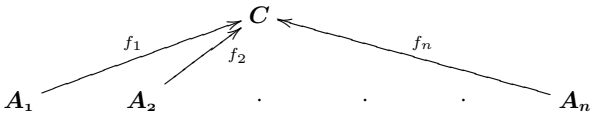
When all the tokens in $Tok(\boldsymbol{A})$ satisfy the $\Vdash_{\boldsymbol{A}}$ relation, then $\Vdash_{\boldsymbol{A}}$ is a subset of $\vdash_{\boldsymbol{A}}$, this latter being the logical consequence, relation.

5. INFORMATION CHANNELS

An *information channel* is a classification used to connect other classifications where the connections are infomorphisms. It is information channels that support information flow by means of *sequents*. An information channel in the binary case (where two classifications are being connected) is a two-way channel. An information channel supports the form of distributed reasoning where one can think of the reasoning as moving along the channel. This is an entirely abstract concept which, given some restrictions, has communication channels as concrete instances.

5.1 Basic Definitions

Definition 5.1.1 (Barwise–Seligman) *An* **information channel** *consists of an indexed family $\mathcal{C} = \{f_i : \boldsymbol{A}_i \longrightarrow \boldsymbol{C}\}$ of infomorphisms with a common codomain $\boldsymbol{C}$ called the* **core** *of the channel. Diagrammatically,*



Frequently in the sequel, the term *channel* will be (mis)used to refer to the core of the channel. This is for mere expediency and the reader is asked to be forgiving. There is never any question as to which morphisms are involved.

Example 5.1.2 Let $\boldsymbol{C}$ model a single user Alice, $\boldsymbol{A}$, sending messages to Eve, $\boldsymbol{E}$. $\boldsymbol{C}$ is to be a channel between $\boldsymbol{A}$ and $\boldsymbol{E}$. Here, a communication channel is being modeled by an information channel. The tokens in the channels are pairs in a relation. Each pair, say $\langle m_1, m_2 \rangle$, is projected by $\overleftarrow{a}$ to Alice's sent message, m_1, and by $\overleftarrow{e}$ to Eve's received message, m_2, and it might be that $m_1 \neq m_2$. The types in each classification below are facts about those mail messages. The channel diagram is

$$\boldsymbol{A} \xrightarrow{\ a\ } \boldsymbol{C} \xleftarrow{\ e\ } \boldsymbol{E}$$

Alice can reason about what Eve knows by reading the mail messages and noticing that the same messages were sent to Eve. Eve can do likewise, hence this is a bi-directional information channel even though the communication channel is from Alice to Eve. In channel theoretic terms, Alice reasons by seeing if a token satisfies sequents of the form $\Gamma \Vdash_{\boldsymbol{A}} \Delta$. and similarly for Eve. However, each uses their local logic in which to do it. To judge the validity of each's reasoning, the local logics can be moved along the infomorphisms via the rules in the previous section.

The above analysis points out that the channels of channel theory are (in general) bidirectional. The reason is they present us with ways of stating properties of the information

of the channel, and those properties are entirely determined by the outside environment, either by ourselves by fiat (convention) or by physical attributes. These properties are then formalized as the types of the channel. The example of current in a wire is a good example. It is only by stipulation that current goes in one direction when in fact it can be looked at as bidirectional for positive and negative charge.

It is helpful to think of a sequent in an information channel as representing a *gate* and to view the classification structure as a mathematical description of the (intuitive view) of a communication channel. The tuples are the information that is produced at $\boldsymbol{A}$, travel through $\boldsymbol{C}$, and arriving at $\boldsymbol{E}$. Each route through $\boldsymbol{C}$ is mediated by a gate (sequent). However, information channels are more general than communication channels. Incidently, the collection of sequents in a channel can model Markov chains, although sequent structure is much more general than Markov structure.

It is $\boldsymbol{A}$'s intention that a fact, say $m^a \models_{\boldsymbol{A}} B$, be communicated to $\boldsymbol{E}$. Assuming no loss of information for the signal, this requires that $\boldsymbol{A}$ and $\boldsymbol{E}$ agree on the types used for communication purposes. The sense of the communication is then

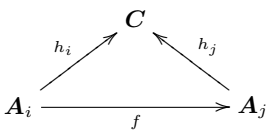
$m^a \models_{\boldsymbol{A}} B$	iff	$m \models_C B^a$	infomorphism condition
	implies	$m \models_C B^e$	?
	iff	$m^e \models_{\boldsymbol{E}} B$	infomorphism condition

where ? indicates a missing reason, namely a sequent of the form $B^a \Vdash_C B^e$. Suppose there are no channel sequents. It is possible for $x^e \models_{\boldsymbol{E}} B$. One could hardly say that communication has taken place because $x^e \models_{\boldsymbol{E}} B$ would have no connection with $x^a \models_{\boldsymbol{E}} B$. In this case, relationship $x^e \models_{\boldsymbol{E}} B$ is spurious or accidental and $\boldsymbol{E}$ can get no information about $\boldsymbol{A}$ from it. The reason probabilities crop up is that $B^a \Vdash_{\boldsymbol{C}} B^e$ may only be partially satisfied, i.e., only some of Alice's messages that satisfy B^a also satisfy B^e.

Definition 5.1.3 (Barwise-Seligman) *A **distributed system** $\mathcal{A}$ consists of an indexed family $cla(\mathcal{A}) = \{\boldsymbol{A}_i\}_{i\in I}$ of classifications together with a set $inf(\mathcal{A})$ of infomorphisms all having both domain and codomain in $cla(\mathcal{A})$.*

A distributed system is simply a collection of classifications and some infomorphisms between some of the classifications. From Barr in [4] reporting on the work of his graduate student Chu, it is clear that categories of classifications have colimits. A colimit of a distributed system is a minimal channel amongst all the channels, each channel connecting the entire distributed system. To be a channel for a distributed system is to *cover* the system. An analogous concept in partial orders is that of an upper bound (think of classifications as points and infomorphisms as elements of the partial order relation), a colimit would be a least upper bound.

Definition 5.1.4 (Barwise-Seligman) *The channel $\mathcal{C} = \{h_i : \boldsymbol{A}_i \longrightarrow \boldsymbol{C}\}_{i\in I}$ **covers** a distributed system $\mathcal{A}$ if for each $i, j \in I$, and each infomorphism $f : \boldsymbol{A}_i \longrightarrow \boldsymbol{A}_j$ in $inf(\mathcal{A})$, the following diagram commutes:*



*$\mathcal{C}$ is a **minimal cover** of a distributed system $\mathcal{A}$ if it covers $\mathcal{A}$ and, for every other channel $\mathcal{D}$ (with core $\boldsymbol{D}$) covering $\mathcal{A}$, there is a unique infomorphism from $\mathcal{C}$ to $\mathcal{D}$.*

Theorem 5.1.5 (Chu [4]) *Every distributed system has a minimal cover, and it is unique up to isomorphism.*

6. PROBABILITY

The probability theory, due to space limitations, will only be worked out for case of CCPL. It provides a template for how the probability theory works out using other logics that have associated measure theories.

There are two ways to view probabilities, either logically inspired as in [12] or set-theoretical as in Kolmogorov [7]. However, our new way to view the situation is to treat the logically inspired axioms as syntax which is then interpreted by the set-theoretical probability functions.

6.1 Basic Probabilities

Probability axioms for CCPL generally do not follow the usual axiomizations of CCPL. Instead, the probability axioms are tuned to picking out intuitive collections of probability functions and the embedding of the axioms within a theory of real numbers allows their extension to all the formulas of CCPL paying mind to the theorems of CCPL. In particular, the following condition is to be proven as a theorem:

AK4: A and B logically equivalent implies $\mathsf{P}(A) = \mathsf{P}(B)$.

The following axioms [12] do allow one to prove **AK4** as a theorem. Consequently, from the soundness and completeness of CPL, all of the theorems of CPL will evaluate under P to 1.

AP1 $0 \leq \mathsf{P}(A)$

AP2 $\mathsf{P}(\neg(A \wedge \neg A)) = 1$

AP3 $\mathsf{P}(A) = \mathsf{P}(A \wedge B) + \mathsf{P}(A \wedge \neg B)$

AP4 $\mathsf{P}(A \wedge B) \leq \mathsf{P}(B \wedge A)$

AP5 $\mathsf{P}(A) \leq \mathsf{P}(A \wedge A)$

One insight of the current paper the recognition that the above axioms (embedded in a real number theory) can be seen as a syntactic system that supports an interpretation by Kolmogorov set functions.

Definition 6.1.1 (Kolmogorov [7]) *Let E be a collection of elements $\xi, \eta, \zeta, \ldots$ which are called **elementary events**, and $\mathfrak{F}$ a set of subsets of E; the elements of the set $\mathfrak{F}$ will be called **random events**.*

- K1. *$\mathfrak{F}$ is a field of sets.*
- K2. *$\mathfrak{F}$ contains the set E.*
- K3. *To each set A in $\mathfrak{F}$ is assigned a non-negative real number $\mathsf{K}(A)$. This number $\mathsf{K}(A)$ is called the probability of the event A.*
- K4. *$\mathsf{K}(E)$ equals 1.*
- K5. *If A and B have no element in common, then*

$$\mathsf{K}(A \cup B) = \mathsf{K}(A) + \mathsf{K}(B).$$

The interpretation first needs a definition (in terms of classifications) for when two types are disjoint:

Definition 6.1.2 *Given a classification $\boldsymbol{A}$, a set $\Gamma \subseteq Typ(\boldsymbol{A})$ is called* **disjoint** *just when for any two types $A, B \in \Gamma$, $Tok(A) \cap Tok(B) = \emptyset$.*

Definition 6.1.3 (Kolmogorov Interpretation) *Given any simple Boolean logic classification, $\boldsymbol{M}$, a Kolomogorov interpretation K is such that*

KI1 $K(0) = 0$;

KI2 $K(1) = 1$;

KI3 $K(\mathsf{P}) = \mathsf{Kr}$ *for* $\mathsf{Kr} : Typ(\boldsymbol{M}) \longrightarrow \mathscr{R}$ *(where $\mathscr{R}$ is the real numbers) such that there is a function* K *satisfying* $\mathsf{K1}$ - $\mathsf{K5}$ *and* $\mathsf{Kr}(A) = \mathsf{K}(Tok(A))$;

KI4 K *takes* $=$ *into* $= \subseteq \mathscr{R} \times \mathscr{R}$;

KI5 K *takes* $\leq$ *into* $\leq \subseteq \mathscr{R} \times \mathscr{R}$.

The following theorem holds:

Theorem 6.1.4 (Soundness) *Given any simple Boolean logic classification, $\boldsymbol{M}$, let K be a Kolmogorov Interpretation function, then $K(\mathsf{P})$ satisfies the $\boldsymbol{AP}$ axioms. The real number theory needed to state and prove theorems using $\boldsymbol{AP}$ axioms is interpreted by the real numbers between* 0 *and* 1.

The only missing element so far is infinite summation. The required logical axiom for this is from [12]:

AP6 $\mathsf{P}(\bigwedge \mathcal{A} \wedge B) = \inf_{Z \subseteq \mathcal{A}} \{\mathsf{P}(B \wedge \bigwedge Z) \mid Z \text{ is finite}\}$

where $\mathcal{A}$ is a countable set, $\bigwedge \mathcal{A}$ is the conjunction of all the elements of $\mathcal{A}$ and similarly for Z. The B is necessary so that the conjunction is never empty.

The associated Kolmogorov axiom is

K6 For a decreasing sequence of events $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ of (the field) $\mathfrak{F}$,

$$\bigcap_{i=1}^{\infty} A_i = \emptyset \text{ implies } \lim_{i \to \infty} \mathsf{K}(A_i) = 0.$$

Soundness is preserved with these additional axioms.

6.2 Conditional Probabilities

Restriction maps are instances of certain quotient morphisms where the relation on types is the identity relation. The tokens of the quotient are not any collection of tokens but rather $Tok(\Gamma)$ for some conjunctive set of types Γ.

Definition 6.2.1 *A* **restriction map** *is an infomorphism $h : \boldsymbol{A} \longrightarrow (\boldsymbol{A} \mid \Gamma)$ such that*

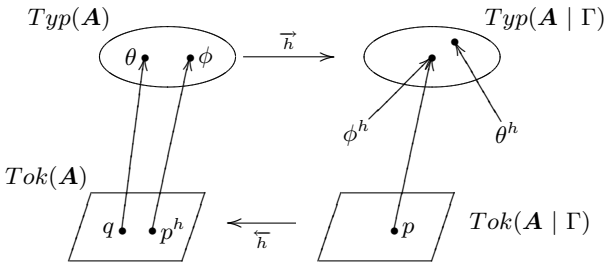
(i): $Typ(\boldsymbol{A} \mid \Gamma) = Typ(\boldsymbol{A})$;

(ii): $Tok(\boldsymbol{A} \mid \Gamma) = Tok(\Gamma)$;

(iii): $\overleftarrow{h}$ *is the injection induced by* $Tok(\Gamma) \subseteq Tok(\boldsymbol{A})$;

(iv): $\overrightarrow{h} = \overrightarrow{1}_{\boldsymbol{A}}$.

The situation can be pictured by:



The (dual) quotient classification $\boldsymbol{A} \mid \Gamma$ will support a probability function that is now viewed as the conditional probability with respect to the types in Γ. There is a condition Mateus, et. al. [8] which must be met for probabilities to be preserved by an infomorphism which promotes the infomorphism to a *probability presentation morphism* Let $h : \boldsymbol{A} \longrightarrow (\boldsymbol{A} \mid \Gamma)$ be a restriction map and let Kr' be the probability function of $\boldsymbol{A}$ and Kr be the conditional probability of $(\boldsymbol{A} \mid \Gamma)$. The condition that must be satisfied is:

$$\mathsf{Kr}'(\bigwedge \Gamma) \times \mathsf{Kr}(\bigvee \Delta \mid \bigwedge \Gamma) = \mathsf{Kr}'(\bigwedge \Gamma \wedge \bigvee \Delta).$$

The restriction that no classical logic classification have an empty set of tokens ensures that $\mathsf{Kr}(\bigvee \Delta \mid \bigwedge \Gamma)$ is never evaluated when $Tok(\bigwedge \Gamma) = \emptyset$.

6.3 Probabilities and Sequents

Probabilities can be assigned to sequents. Consider the simple sequent in $\boldsymbol{A}$ and its satisfying condition:

$$P \Vdash_{\boldsymbol{A}} Q \qquad \forall x \ (x \models_{\boldsymbol{A}} P \textit{ implies } x \models_{\boldsymbol{A}} Q).$$

To attach a probability to this sequent means to weaken it so that it only only holds for some of the tokens and fails to hold the rest. Hence, to weaken the sequent is to remove the universal quantifier and then attach a probability to $x \models_{\boldsymbol{A}} Q$ given that $x \models_{\boldsymbol{A}} P$ for arbitrary x. What is the probability that x satisfies Q given that it satisfies P? This is a statement of conditional probability, so we make the following definition

$$P \Vdash^{\mathsf{P}}_{\boldsymbol{A}} Q \stackrel{def}{=} \mathsf{P}(Q \mid P).$$

To actually use $P \Vdash_{\boldsymbol{A}} Q$ in an argument, one must first have $x \models_{\boldsymbol{A}} P$. The probability of this obtaining in $\boldsymbol{A}$ is $\mathsf{P}(P)$. The use of the rule has the computed probability,

$$\mathsf{P}(P) \cdot (P \Vdash^{\mathsf{P}}_{\boldsymbol{A}} Q).$$

The use of conditional probability to interpret $\Vdash$ is similar to the use of conditional probability in [1] to interpret $\Rightarrow$. In that book, the use of $\Rightarrow$ is derived from conditional probability. Here, the $\Vdash$ is a pre-existing concept which, given a probabilistic clothing, is a definition of conditional probability. This points out that $\Vdash$ is not the same as the material conditional of classical logic and in fact, has no proof theoretic character in channel theory unless provided with a supporting cast which includes a formal system.

Channel theory has sequents of the form $\Gamma \Vdash_{\boldsymbol{A}} \Delta$ for a classification $\boldsymbol{A}$. To use a sequent of this form, P will need to be extended to cover the case of sequents for the following calculation:

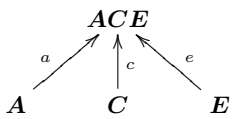
$$\mathsf{P}(\Gamma) \cdot (\Gamma \Vdash^{\mathsf{P}}_{\boldsymbol{A}} \Delta).$$

For a token to satisfy Γ, it must satisfy every element of Γ and hence Γ is thought of conjunctively.

Example 6.3.1 (Capacity Issues) We now study an example from [10], where a standard Shannon-type analysis was done of a covert channel. We show how our new framework extends the classical analysis. There are two users, Alice and Clueless, inside of a private enclave. Alice and Clueless have no knowledge of what the other is doing. The users may transmit no message or one message per unit time to a second enclave. The transmissions between enclaves are encrypted and all messages appear the same to an eavesdropper Eve. The only thing that Eve can do is count the number of messages (per unit time) going from the first enclave (that of Alice and Clueless) to the second enclave. Therefore Eve sees zero, one, or two messages per unit time. Alice uses this scenario to covertly communicate with Eve. Alice will attempt to send a bit to Eve per unit time interval. This is the most that Alice can send because Alice only has two actions. The actions of Clueless act as noise in the covert channel.

Alice will send a 0 by not sending a message. If Alice sends a 0 and Clueless does not transmit, then Eve receives a 0. Alice will send a 1 by sending a message. If Alice sends a 1 and Clueless does not transmit, then Eve receives a 1. If Alice sends a 1 and Clueless does transmit, then Eve receives a 2. Therefore, Eve is only certain of Alice's transmission if Eve receives a 0 or a 2. The received symbol 1 is a noisy symbol. In the following matrix, x_1 represents the actions of Alice, x_2 the actions of Clueless, and x_3 the symbols that Eve receives. The time is in discrete, integral ticks.

Consider the following classification diagram (on the left) of the communication channel



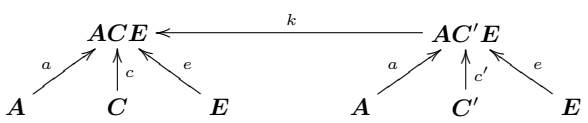
$$\begin{array}{c} \\ 0 \\ 1 \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \end{array}$$

Tokens in the channel are of the form $\langle x_1, x_2, x_3\rangle$ which the allowable values of the combinations of x_i. For $x = \langle x_1, x_2, x_3\rangle$, $\overleftarrow{a}(x) = \langle x_1\rangle$, $\overleftarrow{c}(x) = \langle x_2\rangle$, and $\overleftarrow{e}(x) = \langle x_3\rangle$. Types for component classifications $\boldsymbol{A}$ and $\boldsymbol{C}$ are $\{0, 1\}$ and the types for $\boldsymbol{E}$ are $\{0, 1, 2\}$. These types are injected into the channel (where the superscript indicates which infomorphism did the injection). The channel gates, labeled with g_i, and their respective conditional probabilities are the following:

$$g_1\colon 0^a, 0^c \Vdash_{\boldsymbol{ACE}} 0^e \quad g_2\colon 0^a, 1^c \Vdash_{\boldsymbol{ACE}} 1^e \quad \mathsf{P}(g_1) = \mathsf{P}(g_2) = 1$$

$$g_3\colon 1^a, 0^c \Vdash_{\boldsymbol{ACE}} 1^e \quad g_4\colon 1^a, 1^c \Vdash_{\boldsymbol{ACE}} 2^e \quad \mathsf{P}(g_3) = \mathsf{P}(g_4) = 1$$

Each gate transfers information with probability 1. That is, for every token in the channel, if the left hand side of the gate is satisfied, the right hand side is satisfied. The channel connecting $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{E}$ is taken from a global perspective. To model the system from the more local perspective of only Alice and Eve, the types injected by Clueless must be ignored. Consider an infomorphism k from a new channel to $\boldsymbol{ACE}$:



where $\boldsymbol{C}'$ is has lost the types 0 and 1 and unable to inject them into the channel $\boldsymbol{AC'E}$. The morphism k is stipulated to be the identity on $Tok(\boldsymbol{ACE})$ and an injection on $Typ(\boldsymbol{AC'E})$.

Consider the following use of the second form of k-Elim

$$\frac{0^a, 0^c \Vdash_{\boldsymbol{ACE}} 0^e}{0^a \Vdash_{\boldsymbol{AC'E}} 0^e}\ k\text{-Elim}$$

The conclusion of the rule does not hold because a token of the form $\langle 0, 1, 1\rangle$ is a counter-example to the conclusion whereas the premise is a valid gate in $\boldsymbol{ACE}$. The normal token $\langle 0, 0, 0\rangle$ of $\boldsymbol{ACE}$ will hold of the conclusion, however this cannot be considered a normal token of $\boldsymbol{AC'E}$ since it is a counter-example to the conclusion of another use of k-Elim (see g_2' below). It is but a short step to assign a probability to the conclusions of the four uses of this rule, namely the gates on left below and summarized compactly in a *channel matrix* (identical to that shown in [10]) on the right:

$$g_1'\colon\ 0^a \Vdash^p_{\boldsymbol{AC'E}} 0^e \qquad g_2'\colon\ 0^a \Vdash^q_{\boldsymbol{AC'E}} 1^e$$

$$g_3'\colon\ 1^a \Vdash^\alpha_{\boldsymbol{AC'E}} 1^e \qquad g_4'\colon\ 1^a \Vdash^\beta_{\boldsymbol{AC'E}} 2^e$$

$$\begin{array}{c} \\ 0^a \\ 1^a \end{array} \begin{array}{c} \begin{array}{ccc} 0^e & 1^e & 2^e \end{array} \\ \begin{pmatrix} p & q & 0 \\ 0 & \alpha & \beta \end{pmatrix} \end{array}$$

by using the proportion of tokens which are normal (for each gate alone) to the total number of normal and non-normal tokens (for each gate alone). Incidently, in [10], it is shown that $p = \alpha$ and $q = \beta$ for this example due to the way Clueless acts.

7. CONFIDENTIALITY

This example is modified from [5] and shows how one would go about providing a confidentiality analysis. This example too is much too brief to show all of the new theory's capabilities.

7.1 BLP

Consider Alice and Eve. The following diagram of the channel is used again, but the pieces now contain much different information. In particular, one wishes to reason about properties of messages.

$$\boldsymbol{A} \xrightarrow{\ a\ } \boldsymbol{C} \xleftarrow{\ e\ } \boldsymbol{E}$$

Let Alice send messages m_1, m_2, m_3 and Eve receive messages m_4, m_5, m_6. And let there be types $\alpha, \beta, \gamma, \delta$ in $Typ(\boldsymbol{A})$ and $Typ(\boldsymbol{E})$ where α and β refer to some arbitrary input-output properties, and δ refers to some arbitrary property. γ refers to a confidentiality property which, if set on a source message must not be true of the received message, i.e., the channel must massage the message to remove this property. The classification tables are

$\models_{\boldsymbol{A}\mathrm{or}\boldsymbol{E}}$	α	β	γ
m_1	0	1	1
m_2	1	1	1
m_3	1	0	0

$\models_{\boldsymbol{A}\mathrm{or}\boldsymbol{E}}$	α	β	γ
m_4	1	1	1
m_5	1	0	0
m_6	0	1	1

$\models_{\boldsymbol{C}}$	α^a	α^e	β^a	β^e	γ^a	γ^e	δ
c_1	1	1	1	1	1	1	1
c_2	1	1	0	0	1	0	0
c_3	0	1	1	0	0	1	1

transaction c	file copied ($= c^a$)	resulting file ($= c^e$)
c_1	m_1	m_4
c_2	m_2	m_5
c_3	m_3	m_6

The sequents that must be satisfied in the channel are

fidelity sequents	confidentiality sequents
$\alpha^a \Vdash_{\boldsymbol{C}} \alpha^e$	$\gamma^a \Vdash_{\boldsymbol{C}} \neg\gamma^e$
$\beta^a \Vdash_{\boldsymbol{C}} \beta^e$	

Notice that c_3 violates the sequent involving β. So this sequent has a non-unity probability associated with it. The transactions satisfying the confidentiality sequent are c_2 and c_3. c_1 violates the sequent, since $c_1 \models_{\boldsymbol{C}} \gamma^a$ but $c_1 \not\models_{\boldsymbol{C}} \neg\gamma^e$. c_3 appears odd because it satisfies the confidentiality requirement, but it does so spuriously.

The channel $\boldsymbol{C}$ is not simply a wire over which data is flowing. It requires some internal design to implement the confidentiality policy. In this example, it does not do a particularly fine job. Suppose there is another implementation of that channel, say $\boldsymbol{C}'$. With the probability theory, one can run tests, i.e., greatly expand the number of messages and transactions, and judge the results by putting real probabilities (or better, measures) on the information and comparing how $\boldsymbol{C}'$ does with respect to $\boldsymbol{C}$. One could add a gate (as a sequent) $\gamma^a \Vdash_{\boldsymbol{C}} \gamma^e$ and measure the amount of information leaking under this gate.

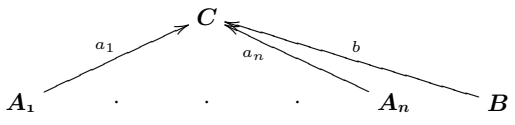
Suppose $\boldsymbol{C}$ is an already existing but faulty channel that cannot be removed except at great expense. Another channel, $\boldsymbol{D}$, could be developed that simply had a better implementation of the confidentiality sequent and interposed between Eve and $\boldsymbol{C}$. The resulting classification structure is

$$\boldsymbol{A} \xrightarrow{a} \boldsymbol{C} \xrightarrow{a} \boldsymbol{D} \xleftarrow{e} \boldsymbol{E}$$

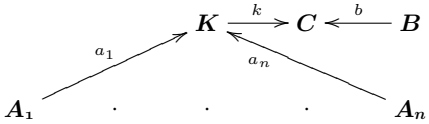
The formalism encourages this sort of composition. The logics in the respective classifications can be moved using the sequent rules. The sequents in the moved logics may acquire new probabilities as a result since not all of those rules preserve validity. This is exactly as it should be because that is what happens in the real world. The theory simply formalizes this so that it cannot be ignored.

7.2 Quasi-Coordinated Attack

Suppose an attack has been made on a system resource $\boldsymbol{B}$ of computer system $\boldsymbol{C}$ possibly involving entities $\boldsymbol{A_1}, \ldots, \boldsymbol{A_n}$. The system has the following initial classification structure:



The classification may contain most but not necessarily all of the known information about the system $\boldsymbol{C}$, the (possibly) attacking entities $\boldsymbol{A_1}, \ldots, \boldsymbol{A_n}$, and the system resource $\boldsymbol{B}$. However, there is no provision here for reconstructing how the attack might have happened. In order to do that, new theories must be developed and compared. One might model this (formally) by constructing a theory for how the attackers conspired. In effect, one needs a theory, i.e., a collection of sequents, which specify how the conspiracy takes place and then map that onto the system to see if it makes sense. One needs the following diagram:



The theory of the conspiracy is now contained in the sequents of $\boldsymbol{K}$. This logical theory (or logic) can be moved using the sequent rules through $\boldsymbol{C}$ to $\boldsymbol{B}$ and now probabilities assigned to see how good a match makes with the observed features of the attack. Those features or observables, in the form of raw data, are the tokens of $\boldsymbol{B}$. The properties the tokens have are the types of $\boldsymbol{B}$. The goodness of the fit will be a measure of how the types (properties of the conspiracy) of $\boldsymbol{K}$ translate through $\boldsymbol{C}$ into types of $\boldsymbol{B}$. The actual artifacts of the conspiracy, i.e., the raw data known about $\boldsymbol{A_1}, \ldots, \boldsymbol{A_n}$, will also get translated with a certain amount of fidelity through $\boldsymbol{C}$ to $\boldsymbol{D}$. All of this can be measured and probabilities assigned so that another conspiracy theory, say $\boldsymbol{K}'$, can also be tried and compared against $\boldsymbol{K}$. Given enough conspiracy theories, the most likely ones (if not all due to limited resources) as measured by the probabilities can be defended against. Not only that, conspiracy theories $\boldsymbol{K_1}, \ldots, \boldsymbol{K_n}$ can be collected together and a meta-theory for this collection produced as a channel connecting them. Collections of meta-theories can be made according to common properties and their features compared using another channel (a meta-meta-theory).

The channel $\boldsymbol{C}$ is probably not something simply explained within one classification. Most likely, it is a whole network of classifications and infomorphisms. This compositionality is handled naturally by the compositionality of this formalism.

8. CONCLUSION

It is clear that our new paradigm has much potential due to its rich qualitative structure and its support of quantitative measures. The qualitative structure does not restrict one to CCPL. There are many special purpose logics that were developed for very particular kinds of problems. Most logics fall into one or another class. Each class indicates a measure theory. The measure theories can then be used to construct Shannon-type information theories. This yields a tight connection between a logic as a qualitative information theory, its associated measure theory, and subsequently, its quantitative information theory.

It is not necessary to use the same logical theory throughout one's analysis. Our paradigm encourages the use of a logic wherever it is appropriate, and the use of several logics (within classifications) connected via infomorphisms provides the glue necessary to get a complete problem analysis.

Our paradigm also supports both top-down and bottom up analyses. The compositionality is neutral in this respect. The compositionality also allows a problem to broken down and assigned to different groups for analyses. And it supports combining different analyses of the same problem by the construction of an information channel supporting a comparison theory of the different analyses.

9. REFERENCES

[1] E. W. Adams. *A Primer of Probability Logic.* CSLI Publications, 1998.

[2] G. Allwein. Probability theory for classifications i: the classical case. Technical report, U.S. Naval Research Laboratory, 2004.

[3] G. Allwein. Probability theory for classifications ii: the general case. Technical report, U.S. Naval Research Laboratory, 2004.

[4] M. Barr. **-Autonomous Categories*. Springer-Verlag, 1979. Lecture Notes in Mathematics 752.

[5] J. Barwise and J. Seligman. *Information Flow: The Logic of Distributed Systems*. Cambridge University Press, 1997. Cambridge Tracts in Theoretical Computer Science 44.

[6] F. I. Dretske. *Knowledge and the Flow of Information*. CSLI Publications, 1999.

[7] A. Kolmogorov. *Foundations of Probability*. Chelsea Publishing Company, New York, 1956. Second English Edition.

[8] P. Mateus, A. Sernadas, and C. Sernadas. Precategories for combining probabilistic automata. *Electronic Notes in Theoretical Computer Science*, 29, 1999.

[9] I. S. Moskowitz, L. Chang, and R. E. Newman. Capacity is the wrong paradigm. In *Proc. New Security Paradigms Workshop, Sept. 23-26*, pages 114–126. ACM Press, 2002.

[10] I. S. Moskowitz, R. E. Newman, D. P. Crepeau, and A. R.Miller. Covert channels and anonymizing networks. In *Proceedings of WPES 2003*, pages 79–93. ACM Press, 2003.

[11] R. Nelson. What is a secret and what does that have to do with computer security. In *Proceedings NSPW, Rhode Island*, pages 74–79, 1994.

[12] P. Roeper and H. Leblanc. *Probability Theory and Probability Logic*. University of Toronto Press, 1999.

[13] C. E. Shannon. The lattice theory of information. *IRE Transactions Information Theory*, (1):180–183, 1950.

[14] C. E. Shannon. *Communication Theory of Secrecy Systems*, pages 84–143. University of Illinois Press, 1993.

[15] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.