A New Framework for Shannon Information Theory

Gerard Allwein and Ira S. Moskowitz and LiWu Chang
Center for High Assurance Computer Systems, Code 5540
Naval Research Laboratory
Washington, D.C. 20375 USA
allwein@itd.nrl.navy.mil, moskowitz@itd.nrl.navy.mil, lchang@itd.nrl.navy.mil

Abstract

Barwise and Seligman proposed a very general qualitative theory of *information flow* (in distributed systems) while Shannon proposed a very general quantitative theory for *communication flow*. The two kinds of flow are not synonymous, with *information flow* being the more general of the two. We synthesize a new theory from these two theories so that the qualitative and quantitative analysis use the same theory structures. The main advantages are (1) Shannon theory gets a more expressive framework within which to operate, (2) Barwise/Seligman theory gets to take advantage of quantitative mechanisms. The resultant theory has direct applications to steganography and covert channels although the development of these applications will appear in a subsequent paper.

1 Introduction

The theory presented in this paper rests upon two particular information theories. The qualitative theory by Barwise and Seligman [3] is known colloquially as *channel theory*. The quantitative theory by Shannon [7] is colloquially known as *information theory*. As several people have noticed (e.g., [4]), Shannon's *information theory* would be better called *communication theory*. We concur and the term *information theory* will be used in the sense of the joint qualitative/quantitative theory we present in this paper. The main difference between the two base theories is how they view *channels*. The Barwise/Seligman notion of *information channel* can be made to support Shannon's notion of *communication channel*. Conversely, Shannon's quantitative methods can provide measures for the Barwise/Seligman notion of channel.

The theory presented here is more general than either Shannon's or Barwise/Seligman's for two reasons:

- Shannon restricted his theory to communication channels. By using quantitative measures on information channels, Shannon's theory is made more inclusive and now applies to this more general notion of channel.
- Barwise/Seligman's theory ignored quantitative measures in favor of a qualitative theory. We make the argument that their qualitative framework can guide a quantitative theory by giving the theory a more expressive scaffolding upon which to apply quantitative measures.

In [5], it is pointed out that the notion of communication channel capacity fails to capture salient features of covert and steganographic channels. In image steganography, information is hidden in a cover image. The Shannon analysis of this situation can put measures on the amount of hidden information the communication channel will support. The problem is that the amounts calculated may have little to do with the transfer of actual information because the information has a qualitative nature to it not amenable to the baseline Shannon analysis. A more sophisticated framework is required upon which to

base the Shannon analysis. Our information theory presented in this paper has a direct application to steganographic analysis in particular and covert channels in general.

2 Classifications and Infomorphisms

The basic unit of information in channel theory is a tuple of a binary relation. The relationship is between a token (a piece of data, say, as in Shannon theory) and a type (what kind of thing is this data of). This is represented as

$$x \models P$$

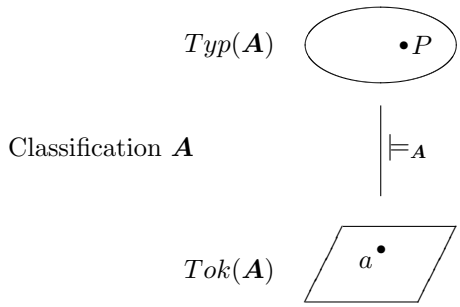
where x is the piece of data, $\models$ is the relation, and P is the type. The symbol, $\models$, is the usual semantic symbol of logic and is usually interpreted in logic as "x satisfies P". This paper will treat $\models$ as the relation "x is of type P". There is to be no metaphysical or epistomological baggage to be associated with "x is of type P" even though we sometimes use the verb "satisfy" when talking about $\models$. Also, one cannot express any property about a single token unless the property is reified as a type and the expression is via the $\models$ relation. Hence, for a number x as a token, one can only express its value V by an expression of the form $x \models V$. In this sense, channel theory enforces a discipline that is sometimes lacking in analysis of information.

To relate the description in the preceding paragraph to Shannon's theory will take most of the work done in the sequel. However, to help orient a reader versed in Shannon's theory, we offer here this description. The basic unit of information in Shannon's theory is also a tuple of a binary relation. The relation is restricted to be of the form $x \models V$ where $\models$ is a function and V is value of the token x. The resulting structure is typically called a state space where V is a state and the tokens are forgotten. Channel theory also has state spaces except the tokens are not forgotten and types are values. States are sometimes further collected together to form events. Channel theory allows this also by first keeping the tokens and then replacing the states as types with events as types. For some event E, $x \models E$ just when $x \models s$ for some state $s \in E$. Hence, Shannon's basic ontology is neatly embedded in channel theory's ontology with channel theory being somewhat more rigorous about the specification of the entities involved.

A collection of types and tokens with their relation is known as a *classification*. A more telling term might be *universe of discourse* and one can freely interchange the two terms. A classification is just what you thought it was, it is a collection of things which have the form of "x is a P", or in our parlance, "x is of type P", i.e., $x \models P$. Information can flow between two classifications via an *infomorphism* which is a special pair of contravariant maps between classifications, one for tokens and one for types. When the information flow between two classifications is of such complexity that it cannot be adequately expressed using a single infomorphism, the flow can be re-expressed as a *channel*. A channel is another classification which is connected to the original two classifications via infomorphisms.

2.1 Classifications

Definition 2.1.1 (Barwise–Seligman) *A* **classification**, $\boldsymbol{A}$, *is a pair of sets and a relation. The sets are called, respectively, the* **tokens**, $Tok(\boldsymbol{A})$, *and* **types**, $Typ(\boldsymbol{A})$. *The binary relation, usually symbolized by* $\models$, *is between the two sets, i.e.,* $\models_{\boldsymbol{A}} \subseteq Tok(\boldsymbol{A}) \times Typ(\boldsymbol{A})$. *The term* $x \models_{\boldsymbol{A}} P$ *means* $\langle x, P\rangle \in \models_{\boldsymbol{A}}$ *with* $x \in Tok(\boldsymbol{A})$ *and* $P \in Typ(\boldsymbol{A})$.



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

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 2.1.2 *Let $\boldsymbol{A} = (Tok(\boldsymbol{A}), Typ(\boldsymbol{A}), \models_{\boldsymbol{A}})$ be a classification, then for any $P \in Typ(\boldsymbol{A})$, $Tok(P) = \{y \mid y \models_{\boldsymbol{A}} P\}$ and, for any $x \in Tok(\boldsymbol{A})$, $Typ(x) = \{Q \mid x \models_{\boldsymbol{A}} Q\}$.*

Example 2.1.3 Let $\boldsymbol{FOL} = (Models, Sentences, \models_{\boldsymbol{FOL}})$ where $Sentences$ are sentences in first order logic (FOL), $Models$ are models of first order sentences, and $x \models_{\boldsymbol{FOL}} S$ iff x is a model of the sentence S. Notice there are a number of internal relations that hold of the set of sentences and the set of models. However, none of these relations are imposed as external conditions in this example. The example could be pumped up to include them. One could also flip this example so that the types were Models and the tokens were Sentences, in which case, Sentences would be classified by Models rather than Models classified by Sentences.

Example 2.1.4 Let $\boldsymbol{T} = (Points, Opens, \models_{\boldsymbol{T}})$ where $Points$ are the points of a topological space, $Opens$ are the set of open sets of that space, and $x \models_{\boldsymbol{T}} O$ iff $x \in O$. This classifies points by the open sets in which they are contained. By reversing the $\models_{\boldsymbol{T}}$, one could classify the opens by the points. The set of open sets forms a Heyting lattice, but that is not specified in this classification and hence no use of this classification within information theory can make use of that fact. It could, however, be imposed on the classification from the outside.

Example 2.1.5 Let $\boldsymbol{M} = (Messages, Contents, \models_{\boldsymbol{M}})$ where Messages are classified by their contents. One could use an entire theory of content in conjunction with the set of types, the theory would have much internal structure. This internal structure is not required by channel theory, but it could be imposed or stipulated if needed.

Example 2.1.6 Let $\boldsymbol{D} = (Times, \{0, 1\}, \models_{\boldsymbol{D}})$ where Times is a set of discrete time stamps, and $t \models_{\boldsymbol{D}} 1$ iff some message was sent at time t and $t \models_{\boldsymbol{D}} 0$ iff no message was sent at time t. Also, $t \not\models_{\boldsymbol{D}} 1$ does not automatically imply $t \models_{\boldsymbol{D}} 0$; there is nothing within channel theory to force this condition. The situation described by this classification might be such that there is incomplete information about whether a message has or has not been sent. You may, however, stipulate (from the outside) such a constraint within the classification. The distinction here from the previous example is that with respect to communication channels, sometimes it is not the messages themselves that are to be modeled but rather information about the messages.

2.2 Infomorphisms

The "flow" of information flow is rarely qualified in many theories of information flow although it is frequently quantified as data flow. Since the currency of information is the tuple "x is of type P", to

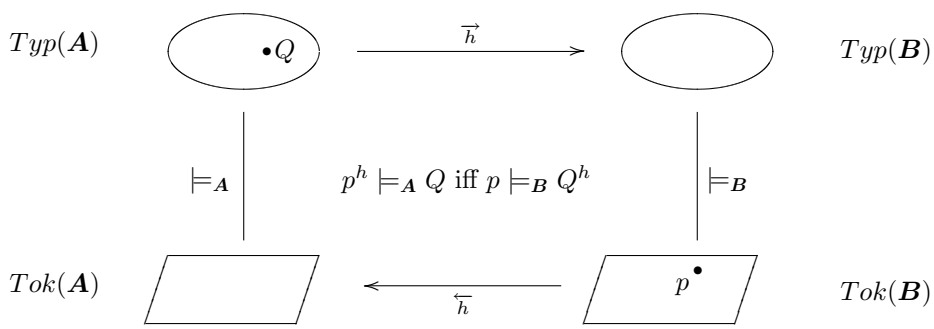
translate information (where here we are using "translate" in its sense as a preservation mapping), one first thinks to translate the x to a y and the P to a Q. This turns out not to be in accord with most uses of classifications within mathematics and logic. More to the point, the morphisms of classifications must relate tokens and types of two classifications in a special way, not simply translate token-type tuples to token-type tuples. The reason for this is that the "flow" of information flow is a flow of logical reasoning, not a flow of the currency.

Definition 2.2.1 (Barwise–Seligman) *Assume classifications* $\boldsymbol{A} = (Tok(\boldsymbol{A}), Typ(\boldsymbol{A}), \models_{\boldsymbol{A}})$ *and* $\boldsymbol{B} = (Tok(\boldsymbol{B}), Typ(\boldsymbol{B}), \models_{\boldsymbol{B}})$. *An* **infomorphism** $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ *is a pair of contravariant maps,* $\overrightarrow{h}$ *and* $\overleftarrow{h}$ *such that* $\overrightarrow{h} : Typ(\boldsymbol{A}) \longrightarrow Typ(\boldsymbol{B})$ *and* $\overleftarrow{h} : Tok(\boldsymbol{B}) \longrightarrow Tok(\boldsymbol{A})$, *and for all* p *and* Q, *the following condition is satisfied:*

$$p^h \models_{\boldsymbol{A}} Q \text{ iff } p \models_{\boldsymbol{B}} Q^h,$$

where for ease of presentation, $\overleftarrow{h}(p)$ *is displayed as* p^h *and* $\overrightarrow{h}(Q)$ *as* Q^h.

This can be pictured with:



The infomorphism h above is (by convention) a morphism from the classification $\boldsymbol{A}$ to the classification $\boldsymbol{B}$. Note that this is not a commutative diagram, the $\models_{\boldsymbol{A}}$ and $\models_{\boldsymbol{B}}$ lines are not arrows or maps. They merely indicate binary relations.

Example 2.2.2 Let $\boldsymbol{SET} = (Models, Sentences, \models_{\boldsymbol{SET}})$ where $Sentences$ are sentences of set theory in the language of FOL, and $Models$ are models of set theory. Let $\boldsymbol{NUM} = (Sentences, Models, \models_{\boldsymbol{NUM}})$ where $Sentences$ are sentences of number theory in the language of FOL and $Models$ are models of number theory. An infomorphism $h : \boldsymbol{NUM} \longrightarrow \boldsymbol{SET}$ might describe number theory as a part of set theory, i.e., translate every sentence in number theory into an equivalent sentence in set theory. The models map goes in the opposite direction, every model of set theory provides a model of number theory. Let m be model of set theory and P some statement of number theory, then

$$m^h \models_{\boldsymbol{NUM}} P \text{ iff } m \models_{\boldsymbol{SET}} P^h$$

says that m^h is model of a sentence P in number theory iff m itself is a model of the interpretation of P, namely P^h, in set theory.

Example 2.2.3 Let $\boldsymbol{T}$ and $\boldsymbol{T}'$ both be topological classifications, then a map $f : Tok(\boldsymbol{T}) \longrightarrow Tok(\boldsymbol{T}')$ is continuous just when f^{-1} is a map from $Typ(\boldsymbol{T}')$ to $Typ(\boldsymbol{T})$. The pair f, f^{-1} constitutes an infomorphism from $\boldsymbol{T}'$ to $\boldsymbol{T}$. For any point x and open set O:

$$x^f \models_{\boldsymbol{T}'} O \text{ iff } x \models_{\boldsymbol{T}} O^f$$

simply because

$$f(x) \in O \text{ iff } x \in f^{-1}(O).$$

Example 2.2.4 Assume there are message classifications $\boldsymbol{M} = (Messages, Contents, \models_{\boldsymbol{M}})$ and $\boldsymbol{M'} = (Messages', Contents', \models_{\boldsymbol{M'}})$. An infomorphism $h : \boldsymbol{M} \longrightarrow \boldsymbol{M'}$ might model a function changing messages from $\boldsymbol{M'}$ to messages in $\boldsymbol{M}$ such that what can be said about the translation can be mapped into something that can be said about the original message:

$$m^h \models_{\boldsymbol{M}} C \text{ iff } m \models_{\boldsymbol{M'}} C^h$$

Here, the translation is working distinctly opposite from that of number theory into set theory.

Example 2.2.5 Let $\boldsymbol{D} = (Times, \{0,1\}, \models_{\boldsymbol{D}})$ and $\boldsymbol{D'} = (Times, \{0,1\}, \models_{\boldsymbol{D'}})$ be two discrete time classifications of messages. An infomorphism $h : \boldsymbol{D} \longrightarrow \boldsymbol{D'}$, defined as $t^h = t'$ where t' means the next time step after t and $N^h = N$ for $N \in \{0,1\}$, models sending messages at one time interval and their reception at the next.

$$t^h \models_{\boldsymbol{D}} 1 \text{ iff } t \models_{\boldsymbol{D'}} 1^h \qquad t^h \models_{\boldsymbol{D}} 0 \text{ iff } t \models_{\boldsymbol{D'}} 0^h$$

says that a message associated with time t^h is received iff a message associated with time t is sent, and that no message is associated with t^h iff no message was sent at time t. Notice that there is no mention that a message sent must be the same as a message received. Again, this is something external to be stipulated. One could easily change the tokens to include the actual messages in order to accommodate this restriction. In this example, the communication channel is modeled as an infomorphism. For more complicated communication channels, this will not be sufficient and the communication channel will be modeled as another classification.

3 Classifications and Probability

3.1 State Spaces and Event Classifications

Definition 3.1.1 (Barwise-Seligman) *A* **state space**, $\boldsymbol{S} = (Tok(\boldsymbol{S}), Typ(\boldsymbol{S}), state_{\boldsymbol{S}})$, *is a classification where* $state_{\boldsymbol{S}} : Tok(\boldsymbol{S}) \longrightarrow Typ(\boldsymbol{S})$ *is a function, i.e., each token is of unique type.*

The tokens are typically abstractions of the system. One might view the tokens as snapshots of the system at various times. The types are typically vectors of values of the system variables. One could reify the tokens as vectors of system variable values at a specified times. In this case, the *state* function merely strips off the time value yielding the vector representing the system state. The reason $state_{\boldsymbol{S}}$ is a function is that a system can be in only one state at a time.

Definition 3.1.2 (Barwise-Seligman) *A* **state space morphism**, $f : \boldsymbol{S_1} \longrightarrow \boldsymbol{S_2}$, *is a pair of maps* $\overleftarrow{f} : Tok(\boldsymbol{S_1}) \longrightarrow Tok(\boldsymbol{S_2})$ *and* $\overrightarrow{f} : Typ(\boldsymbol{S_1}) \longrightarrow Typ(\boldsymbol{S_2})$ *such that*

$$state_{\boldsymbol{S_2}}(x^f) = (state_{\boldsymbol{S_1}}(x))^f.$$

Note both maps run in the same (covariant) direction in contradistinction to infomorphisms which run in opposing (contravariant) directions. Typically, state space analyses totally ignore the notion of token and only the states are deemed important. However, this is not sensitive enough for a qualitative theory where a state may arise for two entirely different reasons. Also, a token is typically not the system itself. Were that the case, there would be only one value or type. A token must be more of an abstract notion of the system at a particular time or place.

Definition 3.1.3 (Barwise-Seligman) *The* **event classification**, $Evt(\boldsymbol{S})$, *associated with a state space* $\boldsymbol{S}$ *has tokens* $Tok(\boldsymbol{S})$ *and types* $Typ(Evt(\boldsymbol{S})) = \mathcal{P}(Typ(\boldsymbol{S}))$ *where* $\mathcal{P}(Typ(\boldsymbol{S}))$ *is the power set of* $Typ(\boldsymbol{S})$.

This definition could be weakened by requiring only $Typ(Evt(\boldsymbol{S})) \subseteq \mathcal{P}(Typ(\boldsymbol{S}))$, but it is handy for Evt to identify a particular event space.

Definition 3.1.4 (Barwise-Seligman) *Given the state space morphism* $f : \boldsymbol{S_1} \longrightarrow \boldsymbol{S_2}$, *the event space morphism* $Evt(f) : Evt(\boldsymbol{S_1}) \longrightarrow Evt(\boldsymbol{S_2})$ *is a pair of contravariant maps where* $\overleftarrow{Evt(f)} = \overleftarrow{f}$ *and* $\overrightarrow{Evt(f)} = \overrightarrow{f}^{-1}$.

3.2 Probability Spaces

The notion of *probability* adheres to the types of a classification. Let $\boldsymbol{A} = (Tok(\boldsymbol{A}), Typ(\boldsymbol{A}), \models_{\boldsymbol{A}})$ be a classification. For E a type, $\mathscr{P}(E)$ is the probability assigned to the type E, and is thought of as the probability that any token is of type E. The probability is taken with respect to the entire set of tokens. Notice that this is distinctly different than the notion of a particular x being of type E with a probability or confidence level of p. This latter might be symbolized with $x \models^p_{\boldsymbol{A}} E$ and represents the idea that $\models_{\boldsymbol{A}}$ (in this instance) is not a concrete relation or a relation we have concrete information about. This has a distinctly Bayesian tinge to it and, although intriguing, we will not consider it here.

Shannon theory always works at the level of types. The reason Shannon theory works at the level of types is because it is working with average amounts of information, not specific pieces of information. Typically, probability theory would force the assumption that the collection of types be a Borel algebra. A probability function $\mathscr{P}$ is then a monotone map from this Borel algebra to the set of real numbers $[0, 1]$, i.e., for x, y members of a Borel algebra upon which $\mathscr{P}$ is defined, $x \leq y$ implies $\mathscr{P}(x) \leq \mathscr{P}(y)$.

There is no structure imposed by channel theory upon any set of types although there is an induced preorder. It is this preorder we take advantage of when defining probability functions for classifications.

Definition 3.2.1 *Given a classification* $\boldsymbol{A}$, *the* **token induced preorder** *on* $Typ(\boldsymbol{A})$ *is defined with*

$$P \prec Q \text{ iff } Tok(P) \subseteq Tok(Q).$$

and the **token induced partial order** *on* $Typ(\boldsymbol{A})$ *is defined with*

$$P \preceq Q \text{ iff } P \prec Q \text{ or } Tok(P) = Tok(Q).$$

The partial order $\preceq$ is essentially $\prec$ divided out by any symmetries induced by equalities of the form $Tok(P) = Tok(Q)$ yielding $P \prec Q$ and $Q \prec P$. The reason to define $\prec$ as a preorder instead of promoting it to a partial order is because the collection of types might have a very intensional description and this would be lost if $\prec$ collapsed types based on the extensional nature of sets alone.

Definition 3.2.2 *Given a classification* $\boldsymbol{A}$, *a set* $\Gamma \subseteq Typ(\boldsymbol{A})$ *is called* **disjoint** *just when for any two types* $P, Q \in \Gamma$, $Tok(P) \cap Tok(Q) = \emptyset$.

Definition 3.2.3 *An* **abstract probability space** *is a classification* $\mathbf{A}$ *together with a probability function* $\mathscr{P} : Typ(\boldsymbol{A}) \longrightarrow Reals$ *satisfying (for types* P *and* Q*):*

($\mathscr{P}$1) for any countable set Γ *of disjoint types with members* P_i,

$$0 \leq \sum_{i=1}^{|\Gamma|} \mathscr{P}(P_i) \leq 1;$$

($\mathscr{P}2$) if $Tok(P) = Tok(\boldsymbol{A})$ then $\mathscr{P}(P) = 1$;
($\mathscr{P}3$) for any countable set Γ of disjoint types with members P_i,

$$[\forall i\ (1 \leq i \leq |\Gamma| \text{ implies } P_i \prec Q)] \text{ implies } \sum_{i=1}^{|\Gamma|} \mathscr{P}(P_i) \leq \mathscr{P}(Q);$$

($\mathscr{P}4$) for any countable set Γ of disjoint types with members P_i,

$$Tok(Q) \subseteq \bigcup (Tok(P_i) \mid P_i \in \Gamma) \text{ implies } \mathscr{P}(Q) \leq \sum_{i=1}^{|\Gamma|} \mathscr{P}(P_i);$$

($\mathscr{P}5$) $Tok(P) = \emptyset$ implies $\mathscr{P}(P) = 0$.

The axiom ($\mathscr{P}2$) is different than in Kolmogorov's axioms for the simple reason that there need not be a type which all tokens satisfy. One can always adjoin a type, U, to the types (of a classification) such that all tokens satisfy U and it will not affect the exposition here. The third axiom implies that $\mathscr{P}$ is monotone and hence $P \prec Q$ and $Q \prec P$ imply $\mathscr{P}(P) = \mathscr{P}(Q)$. The third and the forth axioms will force the Kolmogorov axiom

$$\mathscr{P}(\bigvee_i \{P_i \mid P_i \in \Gamma\}) = \sum_{i=1}^{|\Gamma|} \mathscr{P}(P_i)$$

to be true if the collection of types is a Borel algebra. The last axiom is an abstraction of the situation where event spaces are generated from state spaces and the observation that if a state cannot occur, i.e., it has no tokens, then it must have probability of 0. Events are collections of states, so if the states of an event have no tokens, the event has probability 0.

Theorem 3.2.4 *If $\boldsymbol{A}$ is an abstract probability space when $Typ(\boldsymbol{A})$ under the token induced $\preceq$ partial order has a Borel lattice structure, then $\boldsymbol{A}$ is a probability space.*

proof: The Kolmogorov axioms for a probability space are easily seen to be true under these conditions. ∎

The $\prec$ order is a remnant of the Boolean lattice order of a Borel algebra. In fact, if the classification does arise as an event space from a state space, the $\preceq$ order is very nearly the $\subseteq$ order on the event space. The only difference is that the event space definition does not require a state have tokens yet the $\preceq$ order is defined entirely in terms of tokens. If this is the case, i.e., every state of every event has tokens, then the two orders are isomorphic. In any case, with the exclusion of the last axiom, every probability space defined on an event space is an abstract probability space since the first three axioms are easily satisfied and, since $\cup$ is the least upper bound, the forth axiom is satisfied.

Theorem 3.2.5 *Let $\mathcal{E} = (Typ(Evt(\boldsymbol{S})), \bigcap, \bigcup, -, \top, \bot)$ be the complete Boolean lattice of sets where $Evt(\boldsymbol{S})$ is the event space defined from a state space $\boldsymbol{S}$ and $\top = Typ(\boldsymbol{S})$ and $\bot = \emptyset$. Let*

$$\mathcal{T}(\mathcal{E}) = \{u \mid u = Tok(P) \text{ and } P \in Typ(Evt(\boldsymbol{S}))\}.$$

Then $\mathcal{T} = (\mathcal{T}(\mathcal{E}), \bigcap, \bigcup, -, Tok(\boldsymbol{S}), \emptyset)$ is a complete Boolean lattice of token sets. The function $f : \mathcal{E} \longrightarrow \mathcal{T}$ where

$$f(P) = Tok(P) = \{x \mid x \models_{\boldsymbol{S}} s \text{ and } s \in P\}$$

is a lattice epimorphism sending $\top$ to $Tok(\boldsymbol{S})$ and $\bot$ to $\emptyset$. If $Tok(s) \neq \emptyset$ for all $s \in Typ(\boldsymbol{S})$ then $Tok(-)$ is also 1-1.

proof: Let $\Gamma \subseteq Typ(Evt(\boldsymbol{S}))$ then

$$\begin{aligned} x \in Tok(\bigcup\{P \mid P \in \Gamma\}) \ \textit{iff}\ & (\exists s \ \in \bigcup\{P \mid P \in \Gamma\})[x \models_{\boldsymbol{S}} s] \\ \textit{iff}\ & (\exists s\ )(\exists P \in \Gamma\ )[x \models_{\boldsymbol{S}} s \ \textit{and}\ s \in P] \\ \textit{iff}\ & (\exists P \in \Gamma\ )[x \in Tok(P)] \\ \textit{iff}\ & x \in \bigcup\{Tok(P) \mid P \in \Gamma\} \end{aligned}$$

and

$$\begin{aligned} x \in Tok(\bigcap\{P \mid P \in \Gamma\}) \ \textit{iff}\ & (\exists s \ \in \bigcap\{P \mid P \in \Gamma\})[x \models_{\boldsymbol{S}} s] \\ \textit{iff}\ & (\exists s\ \forall P \in \Gamma\ )[x \models_{\boldsymbol{S}} s \ \textit{and}\ s \in P] \\ \textit{iff}\ & (\forall P \in \Gamma\ )[x \in Tok(P)] \\ \textit{iff}\ & x \in \bigcap\{Tok(P) \mid P \in \Gamma\} \end{aligned}$$

and for $P \in Typ(Evt(\boldsymbol{S}))$,

$$\begin{aligned} x \in Tok(\top - P) \ \textit{iff}\ & (\exists s \ \in \top - P)[x \in Tok(s)] \\ \textit{iff}\ & x \notin Tok(P) \\ \textit{iff}\ & x \in Tok(\boldsymbol{S}) - Tok(P) \end{aligned}$$

and

$$\begin{aligned} x \in Tok(\top) \ \textit{iff}\ & (\exists s \ \in \boldsymbol{S})[x \in Tok(s)] \\ \textit{iff}\ & x \in Tok(\boldsymbol{S}) \end{aligned} \qquad \begin{aligned} x \in Tok(\bot) \ \textit{iff}\ & (\exists s \ \in \bot)[x \in Tok(s)] \\ \textit{iff}\ & x \in \emptyset \end{aligned}$$

This shows the set operations on $\mathcal{T}(\mathcal{E})$ are well-defined and that f is an epimorphism. Assume $Tok(s) \neq \emptyset$ for all $s \in Typ(\boldsymbol{S})$ and for $P, Q \in Typ(Evt(\boldsymbol{S}))$, $P \neq Q$. Without loss of generality, let $s \in P$ and $s \notin Q$. Since $Tok(s) \neq \emptyset$, there is some x such that $x \models_{\boldsymbol{S}} s$ and $x \in Tok(P)$. Since $\models_{\boldsymbol{S}}$ is a function (as opposed to a mere relation), $x \notin Tok(Q)$ and therefore f is 1-1. ∎

Example 3.2.6 Suppose there is a physical system which is described by a state space $\boldsymbol{S}$. Let $Tok(\boldsymbol{S})$ be instances of the system at various times. Time need have no beginning and end for the system although you can impose one. The states of the system are vectors of measurable properties. The event space $Evt(\boldsymbol{S})$ has as types the power set of $Typ(\boldsymbol{S})$, and as tokens $Tok(\boldsymbol{S})$ such that $x \models_{Evt_{\boldsymbol{S}}} E$ iff there is some $s \in E$ and $state_{\boldsymbol{S}}(x) = s$. The proportion of time the system spends in a particular state, s, is modeled as $\mathscr{P}(s)$. The proportion of time associated with an event E is $\mathscr{P}(E)$ which totals up all the time the system spends in any of the states of E.

Incidently, the $\prec$ order turns out to be preserved by infomorphisms, i.e., they are monotone maps on types. That this order is preserved by infomorphisms without any extra conditions shows that this order is an intrinsic feature for this category of classifications.

Theorem 3.2.7 *Let $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be an infomorphism, then $P \prec Q$ implies $P^h \prec Q^h$.*

proof: Assume $P \prec Q$ and let $x \in Tok(P^h)$, then $x \models_{\boldsymbol{B}} P^h$. From the infomorphism condition, $x^h \models_{\boldsymbol{A}} P$ and hence $x^h \in Tok(P)$. Since $P \prec Q$, $x^h \in Tok(Q)$ and $x^h \models_{\boldsymbol{A}} Q$. From the infomorphism condition again, $x \models_{\boldsymbol{B}} Q^h$ and hence $x \in Tok(Q^h)$. By definition $P^h \prec Q^h$. ∎

Dually, one can define a preorder on tokens by extending $Typ(-)$ within a classification $\boldsymbol{A}$ using $Typ(x) \stackrel{def}{=} \{P \mid x \models_{\boldsymbol{A}} P\}$. And a similar theorem will show that $x \prec y$ implies $x^h \prec y^h$ for $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$.

Theorem 3.2.8 *Let $h : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be an infomophism, then $\overleftarrow{h}(Tok(P^h)) \subseteq Tok(P)$ and $\overrightarrow{h}(Typ(x^h)) \subseteq Typ(x)$.*

The proof is a simple application of the infomorphism condition. In fact, these two theorems taken as axioms completely characterize infomorphisms.

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:

$$(\textit{for all } P \in \Gamma, x \models_{\boldsymbol{A}} P) \textit{ implies } (\textit{there exists one } Q \in \Delta, x \models_{\boldsymbol{A}} Q).$$

When Γ or Δ are singleton sets, say, $\{A\}$, then $A \vdash \Delta$ or $\Gamma \vdash A$ 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. To impose structure on the collection of types means that that structure is not accessible via channel infomorphisms (i.e., at the level of category theory) and only accessible at the additional cost of extra mathematical scaffolding. 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** $\mathfrak{L} = \langle \boldsymbol{A}, \vdash_{\mathfrak{L}}, N_{\mathfrak{L}} \rangle$ *consists of a classification $\boldsymbol{A}$, a set $\vdash_{\mathfrak{L}}$ of sequents involving the types of $\boldsymbol{A}$, and a subset $N_{\mathfrak{L}} \subseteq Tok(\boldsymbol{A})$ called the* **normal tokens** *of $\mathfrak{L}$, which satisfy all the constraints $\vdash_{\mathfrak{L}}$. A local logic $\mathfrak{L}$ is sound if every token is normal; it is complete if every sequent that holds of all normal tokens is in the consequence relation $\vdash_{\mathfrak{L}}$.*

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 $f : \boldsymbol{A} \longrightarrow \boldsymbol{B}$:

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

where Γ^{-f} is a nicer way of writing $\overrightarrow{f}^{-1}$, 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 $f(x) \models_{\boldsymbol{A}} P$ 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 $\boldsymbol{A}$.

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

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

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

$$P \vdash_{\boldsymbol{A}}^{\mathscr{P}} Q \stackrel{def}{=} \mathscr{P}(Q \mid P).$$

When a sequent's conditional probability is known to be, say p, then this will be indicated by

$$P \vdash_{\boldsymbol{A}}^{p} Q.$$

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 $\mathscr{P}(P)$. The use of the rule has the computed probability,

$$\mathscr{P}(P) \cdot (P \vdash_{\boldsymbol{A}}^{\mathscr{P}} Q).$$

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, $\mathscr{P}$ will need to be extended to cover the case of sequents for the following calculation:

$$\mathscr{P}(\Gamma) \cdot (\Gamma \vdash_{\boldsymbol{A}}^{\mathscr{P}} \Delta).$$

For a token to satisfy Γ, it must satisfy every element of Γ and hence Γ is thought of conjunctively. In probability theory, this is generally not a problem since Γ could be equated to $\bigwedge \Gamma$ and $Typ(\boldsymbol{A})$ would actually be a Borel lattice of sets with countable meets and joins. Space prevents us from a complete exposition here but the trick is to transfer $\mathscr{P}$ from $Typ(\boldsymbol{A})$ to the tokens set $\{Tok(P) \mid P \in Typ(\boldsymbol{A})\}$. As a set of sets, a semilattice can be generated isomorphic to the free meet semilattice using $Typ(\boldsymbol{A})$ as generators. This is essentially the intersection semilattice generated by $\{Tok(P) \mid P \in Typ(\boldsymbol{A})\}$. A similar construction must be done for the complete join semilattice to evaluate $\mathscr{P}(\Delta \mid \Gamma)$. However, a further quotient join semilattice must be constructed by dividing the join semilattice with a collection of equalities of the form $P = 1$ for all $P \in \Gamma$ where 1 is the top of the join semilattice.

A *good* probability function is any that satisfies the axioms $(\mathscr{P}1)-(\mathscr{P}5)$ now transferred to $\{Tok(P) \mid P \in Typ(\boldsymbol{A})\}$ and is defined over the newly introduced semilattices. Space prevents us from a fuller exposition here and the sequents used in examples in the sequel will be only of the form $P \vdash_{\boldsymbol{A}} Q$ or $P, P' \vdash_{\boldsymbol{A}} Q$.

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*. One might think that the notion of a "channel" should be captured by an infomorphism. 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** *is 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,*

$$\begin{tikzcd}
 & & \boldsymbol{C} & & \\
\boldsymbol{A_1} \arrow[urr, "f_1"] & \boldsymbol{A_2} \arrow[ur, "f_2"] & \cdot & \cdot & \cdot & \boldsymbol{A_n} \arrow[ull, "f_n"']
\end{tikzcd}$$

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 sending messages to two different people, Alice, modeled by $\boldsymbol{A}$ and Eve, modeled by $\boldsymbol{E}$. $\boldsymbol{C}$ is to be a channel between $\boldsymbol{A}$ and $\boldsymbol{E}$ but notice this is not a communication channel since neither Alice nor Eve are sending messages to each other. The tokens are the individual mail messages with Alice and Eve both mentioned as recipients in the message headers. The types are facts about those mail messages. Let $\boldsymbol{A}$, $\boldsymbol{C}$, and $\boldsymbol{E}$ all share the same types and the same tokens. The channel diagram is

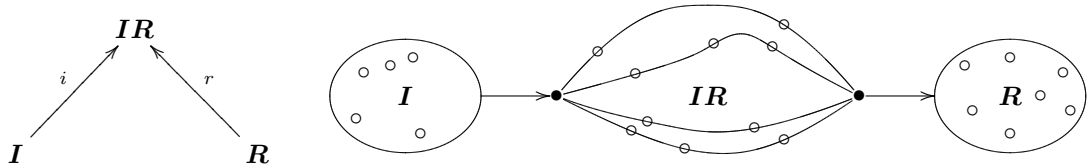
$$\boldsymbol{A} \xrightarrow{id} \boldsymbol{C} \xleftarrow{id} \boldsymbol{E}$$

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. In channel

theoretic terms, Alice reasons by seeing if a token satisfies sequents of the form $\Gamma \vdash_{\boldsymbol{A}} \Delta$. Since all the infomorphisms are the identity morphism, Alice knows that $\Gamma \vdash_{\boldsymbol{C}} \Delta$ holds in the channel and that $\Gamma \vdash_{\boldsymbol{E}} \Delta$ holds for Eve.

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.

Example 5.1.3 Consider the case of an initiator of communication, $\boldsymbol{I}$, and a receiver, $\boldsymbol{R}$ with a binary channel, $\boldsymbol{IR}$, between them meant to represent a communication channel:



Let $\overleftarrow{i} : Tok(\boldsymbol{IR}) \rightarrowtail Tok(\boldsymbol{I})$ and $\overleftarrow{r} : Tok(\boldsymbol{IR}) \rightarrowtail Tok(\boldsymbol{R})$, i.e., $\overleftarrow{i}$ and $\overleftarrow{r}$ are like projection maps except that they inject tokens from the channel into the token sets of $\boldsymbol{I}$ and $\boldsymbol{R}$. The injections model that $\boldsymbol{I}$ only sends part of $Tok(\boldsymbol{I})$ and $\boldsymbol{R}$ only receives part of $Tok(\boldsymbol{R})$.

Let a sequent in an information channel representing a communication channel be called a *gate*. It is tempting to view the classification structure (on the left above) as a mathematical description of the (intuitive view) of a communication channel (on the right) where each $\circ$ in the respective classifications is a tuple of the $\models$ relation specific to that classification. The tuples are the information that is produced at $\boldsymbol{I}$, travel through the $\boldsymbol{IR}$ via one of the routes, and arriving at $\boldsymbol{R}$. Each route is mediated by a gate (sequent) to which a probability will be assigned.

This second diagram is misleading in the sense that information tuples do not actually move in the classification scheme. Instead, there are static mathematical relationships which relate tuples of the classification $\boldsymbol{I}$ to those of $\boldsymbol{IR}$, and similarly, tuples of $\boldsymbol{R}$ to those of $\boldsymbol{IR}$. It is our external claim that the mathematics models the communication channel. Now that there is a mathematical model, however, it can be tested to see with what degree of fidelity it models the real situation.

The initiator $\boldsymbol{I}$ is intending to send not simply a message m^i but instead the tuple $\langle m^i, C\rangle \in \models_{\boldsymbol{I}}$ since this is the basic unit of currency in channel theory. It is the image, m^i, under infomorphism i of channel message m that $\boldsymbol{I}$ is sending. If there is no C for which $m^i \models_{\boldsymbol{I}} C$, then in effect there is nothing $\boldsymbol{I}$ can say about m^i. The communication can still take place, but nothing can be said about actual value transferred.

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

$m^i \models_{\boldsymbol{I}} C$	*iff*	$m \models_{\boldsymbol{IR}} C^i$	infomorphism condition
	implies	$m \models_{\boldsymbol{IR}} C^r$	?
	iff	$m^r \models_{\boldsymbol{R}} C$	infomorphism condition

where ? indicates a missing reason (supplied below). Clearly, this should be the case for all types C. Necessarily, $\boldsymbol{I}$ and $\boldsymbol{R}$ must have agreed on channel sequents for these (but not all) types. Suppose there are no channel sequents. It is possible for $x^r \models_{\boldsymbol{R}} C'$ for some C'. One could hardly say that communication

has taken place because C' has no connection with $\boldsymbol{I}$. The relationship $x^r \models_{\boldsymbol{R}} C'$ is spurious or accidental and $\boldsymbol{R}$ can get no information about $\boldsymbol{I}$ from it.

Note that one cannot even tell in which direction the communication is taking place. To actually say something about the communication, one must classify precisely what is to be said. Let there be types $Src \in Typ(\boldsymbol{I})$ and $Dst \in Typ(\boldsymbol{R})$ such that for all tokens $x \in Tok(\boldsymbol{IR})$,

$$x^i \models_{\boldsymbol{I}} Src \qquad\qquad x^r \models_{\boldsymbol{R}} Dst.$$

Now, the channel models a direction via the stipulated conditions on infomophisms, i.e.,

$$x^i \models_{\boldsymbol{I}} Src \text{ iff } x \models_{\boldsymbol{IR}} Src^i \qquad\qquad x^r \models_{\boldsymbol{R}} Dst \text{ iff } x \models_{\boldsymbol{IR}} Dst^r,$$

and with channel tokens satisfying the following gate on the left via the condition on the right:

$$Src^i \vdash_{\boldsymbol{IR}} Dst^r \qquad\qquad \text{for all } z \in Tok(\boldsymbol{IR}), z \models_{\boldsymbol{IR}} Src^i \text{ implies } z \models_{\boldsymbol{IR}} Dst^r.$$

This gate supplies the missing condition (?) above for $C^i = Src^i$ and $C^r = Dst^r$. This stipulated direction through the channel appears artificial but it is also echoed in information transfer in Shannon's theory. There, all one has is measurement of the information that was transferred. From the measurements alone, it is impossible to tell the direction of information flow.

Definition 5.1.4 (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* $cla(\mathcal{A})$.

A distributed system is simply a collection of classifications and some infomorphisms between some of the classifications. From Barr in [2] 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 5.1.5 (Barwise-Seligman) *A 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:*

$$\begin{tikzcd} & \boldsymbol{C} & \\ \boldsymbol{A}_i \arrow[ur, "h_i"] \arrow[rr, "f"'] & & \boldsymbol{A}_j \arrow[ul, "h_j"'] \end{tikzcd}$$

$\mathcal{C}$ *is a* **minmal 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 5.1.6 (Chu) *Every distributed system has a minimal cover, and it is unique up to isomorphism.*

Example 5.1.7 An infomorphism is a restricted form of a channel. This construction of a channel from an infomorphism is instructive in that it shows how much more freedom there is in the notion of an information channel. Let $f : \boldsymbol{A} \longrightarrow \boldsymbol{B}$ be an infomorphism. The intuitive idea is to represent $\overleftarrow{f}$ in its graph form and $\overrightarrow{f}$ as quotient on the disjoint union of $Typ(\boldsymbol{A})$ and $Typ(\boldsymbol{B})$. The colimit of this as a distributed system has the following intuitive diagram on the left specifying the conditions on the right:

$$\begin{tikzcd} & \boldsymbol{C} & \\ \boldsymbol{A} \arrow[ur, "\pi_1"] \arrow[rr, "f"'] & & \boldsymbol{B} \arrow[ul, "\pi_2"'] \end{tikzcd}$$

$$\overleftarrow{f}(x) = \overleftarrow{f}(\overleftarrow{\pi}_2(\langle y, x\rangle)) = \overleftarrow{\pi}_1(\langle y, x\rangle) = y,$$

$$\overrightarrow{\pi}_1(\overrightarrow{f}(Q)) = \overrightarrow{\pi}_2(Q).$$

for all types Q. This simply assures a type Q from $\boldsymbol{A}$ is sent to the same type as $\overrightarrow{f}(Q)$ when both are injected into the channel $\boldsymbol{C}$.

5.2 Modeling Communication

We now study an example from [6], where a standard Shannon-type analysis was done of a covert channel. We show how our new framework extends the classical analysis. The scenario is simple: 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

$$\begin{tikzcd} & \boldsymbol{ACE} & \\ \boldsymbol{A} \arrow[ur, "a"] & \boldsymbol{C} \arrow[u, "c"] & \boldsymbol{E} \arrow[ul, "e"'] \end{tikzcd} \qquad \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & t \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & t_1 \\ 1 & 0 & 1 & t_2 \\ 0 & 1 & 1 & t_3 \\ 1 & 1 & 2 & t_4 \end{pmatrix} \end{array}$$

Tokens in the channel are of the form $\langle x_1, x_2, x_3, t\rangle$ where the allowable values of the combinations of x_i and tick time in the tuples are recorded in the matrix above (t is a natural number). For $x = \langle x_1, x_2, x_3, t\rangle$, $\overleftarrow{a}(x) = \langle x_1, t\rangle$, $\overleftarrow{c}(x) = \langle x_2, t\rangle$, and $\overleftarrow{e}(x) = \langle x_3, t\rangle$. 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:

$$g_1\colon\ 0^a, 0^c \vdash_{\boldsymbol{ACE}} 0^e \quad g_2\colon\ 0^a, 1^c \vdash_{\boldsymbol{ACE}} 1^e \qquad \mathscr{P}(0^e \mid 0^a, 0^c) = 1 \quad \mathscr{P}(1^e \mid 0^a, 1^c) = 1$$

$$g_3\colon\ 1^a, 0^c \vdash_{\boldsymbol{ACE}} 1^e \quad g_4\colon\ 1^a, 1^c \vdash_{\boldsymbol{ACE}} 2^e \qquad \mathscr{P}(1^e \mid 1^a, 0^c) = 1 \quad \mathscr{P}(2^e \mid 1^a, 1^c) = 1$$

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 $\boldsymbol{ACE}$:

$$\begin{tikzcd}
 & \boldsymbol{ACE} & & & \boldsymbol{AC'E} & \\
\boldsymbol{A} \arrow[ur, "a"] & \boldsymbol{C} \arrow[u, "c"] & \boldsymbol{E} \arrow[ul, "e"'] & \boldsymbol{A} \arrow[ur, "a"] \arrow[rr, phantom, "\longleftarrow k \longleftarrow"] & \boldsymbol{C'} \arrow[u, "c'"] & \boldsymbol{E} \arrow[ul, "e"']
\end{tikzcd}$$

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 $Typ(\boldsymbol{AC'E})$. In general, for the infomorphism $f : \boldsymbol{X} \longrightarrow \boldsymbol{Y}$, the rules

$$\frac{\Gamma^f \vdash_{\boldsymbol{Y}} \Delta^f}{\Gamma \vdash_{\boldsymbol{X}} \Delta}\ f\text{-Elim} \qquad\qquad \frac{\Gamma \vdash_{\boldsymbol{Y}} \Delta}{\Gamma^{-f} \vdash_{\boldsymbol{X}} \Delta^{-f}}\ f\text{-Elim}$$

do not preserve validity (as previously noted). However, they fail to do so for very different reasons. The first form is from Barwise/Seligman [3], the second form is new. The first fails because there can easily be tokens of $\boldsymbol{X}$ which fail the conclusion, but they will not be of the form $f(x)$ for $x \in Tok(\boldsymbol{Y})$. The second form can fail because not every type of $\boldsymbol{Y}$ need be a type of $\boldsymbol{X}$. Hence, even if $Tok(\boldsymbol{X}) = Tok(\boldsymbol{Y})$, tokens that inadvertently satisfied $\Gamma \vdash_{\boldsymbol{Y}} \Delta$ by failing to satisfy all types in Γ might easily satisfy all types in Γ^{-f} simply because the use of sets and functions only guarantee that $(\Gamma^{-f})^f \subseteq \Gamma$, not $(\Gamma^{-f})^f = \Gamma$

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 $\langle 0, 1, 1, t\rangle$ is a counter-example to the conclusion whereas the premise is a valid gate in $\boldsymbol{ACE}$. The normal token $\langle 0, 0, 0, t\rangle$ 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 [6]) on the right:

$$g_1'\colon\ 0^a \vdash^p_{\boldsymbol{AC'E}} 0^e \qquad g_2'\colon\ 0^a \vdash^q_{\boldsymbol{AC'E}} 1^e \qquad\qquad \begin{array}{c} \\ 0^a \\ 1^a \end{array} \begin{array}{c} \begin{array}{ccc} 0^e & 1^e & 2^e \end{array} \\ \begin{pmatrix} p & q & 0 \\ 0 & \alpha & \beta \end{pmatrix} \end{array}$$

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

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 [6], it is shown that $p = \alpha$ and $q = \beta$ for this example due to the way Clueless acts.

5.3 Modeling Diagram Transmission

There is an interesting twist on the previous example which was brought up in [5] in connection with steganography . The details are changed here due to space consideration. Suppose a bitmap of a picture is to be sent and the picture contains the diagram of the numeral 1. Suppose further that there is a noise producer similar to the previous example in that it flips bits in according to a uniform random distribution. The question is, how is it that even with moderate amounts of noise and reduced channel capacity, the 1 is still able to be received and recognized as a 1. Informationally, the noise, unless it rises high enough, does nothing to degrade the information being sent.

There are several different ways of modeling the situation in channel theory. Essentially, they all reduce to there being another channel involved that is derived from and in addition to the existing communication channel. Specifically, assume the channel $\boldsymbol{ACE}$ from the previous example, except that here Alice $\boldsymbol{A}$ is now sending the bits of the picture, $\boldsymbol{C}$ is the noise producer and Eve $\boldsymbol{E}$ is receiving the picture. There is another channel with gates:

$$\begin{tikzcd} & \widehat{\boldsymbol{ACE}} & \\ \widehat{\boldsymbol{A}} \arrow[ur, "\hat{a}"] & \widehat{\boldsymbol{C}} \arrow[u, "\hat{c}"] & \widehat{\boldsymbol{E}} \arrow[ul, "\hat{e}"'] \end{tikzcd}$$

$$\hat{g}_1\colon\ \mathbb{O}^{\hat{a}}, \mathbb{F}^{\hat{c}} \vdash_{\widehat{\boldsymbol{ACE}}} \mathbb{O}^{\hat{e}} \qquad \hat{g}_2\colon\ \mathbb{I}^{\hat{a}}, \mathbb{F}^{\hat{c}} \vdash_{\widehat{\boldsymbol{ACE}}} \mathbb{I}^{\hat{e}}$$

$$\hat{g}_3\colon\ \mathbb{O}^{\hat{a}}, \mathbb{F}^{\hat{c}} \vdash_{\widehat{\boldsymbol{ACE}}} \mathbb{I}^{\hat{e}} \qquad \hat{g}_3\colon\ \mathbb{I}^{a}, \mathbb{F}^{\hat{c}} \vdash_{\widehat{\boldsymbol{ACE}}} \mathbb{O}^{\hat{e}}$$

This is a derived channel where $Tok(\widehat{\boldsymbol{A}}) = \mathcal{P}(Tok(\boldsymbol{A}))$, $Tok(\widehat{\boldsymbol{E}}) = \mathcal{P}(Tok(\boldsymbol{E}))$, and $Tok(\widehat{\boldsymbol{C}}) = \{f\}$ for f the noise producing function defined such that for token $\langle X, f, Y\rangle$ with $X \in Tok(\widehat{\boldsymbol{A}})$ and $Y \in Tok(\widehat{\boldsymbol{E}})$, $f_n(X) = Y$. In short, the token sets are up one set theoretical type level from the token sets of the originating classification. The types for $\widehat{\boldsymbol{A}}$ and $\widehat{\boldsymbol{E}}$ are $\mathbb{I}$ for the diagram of 1 and $\mathbb{O}$ for no diagram of 1. Similarly to the previous example, $\mathbb{I}^{\hat{a}}$ represents the type $\mathbb{I}$ injected into the channel from $Typ(\widehat{\boldsymbol{A}})$. The lone type of $\widehat{\boldsymbol{C}}$ is $\mathbb{F}$.

For classification $\widehat{\boldsymbol{A}}$, let $X \models_{\widehat{\boldsymbol{A}}} \mathbb{I}$ just when X appears as a picture of the diagram of 1 and $X \models_{\widehat{\boldsymbol{A}}} \mathbb{O}$ otherwise. X necessarily includes some surrounding pixels so that that the 1 may be discerned from the background. The situation is similar for $\widehat{\boldsymbol{E}}$. For noise producer $\boldsymbol{C}$, $f_n \models_{\widehat{\boldsymbol{C}}} \mathbb{F}$.

Every token in the channel satisfies one of the gates with the interpretation that the token $\mathbb{I}^{\hat{a}}$ just when Alice thinks it looks like a 1 and satisfies $\mathbb{I}^{\hat{e}}$ just when Eve thinks it looks like a 1. The token satisfies the respective $\mathbb{O}$ types otherwise. Every token satisfies $\mathbb{F}^{\hat{c}}$ because the image of every token in $Tok(\widehat{\boldsymbol{C}})$ under the infomorphism $\hat{c}$ satisfies $\mathbb{F}$ in $\boldsymbol{C}$.

The net result is that $\mathbb{F}^{\hat{c}}$ is parametric to the gates since every token satisfies it and it appears in every gate. Hence forming the infomorphism $\hat{k}$, similar to k in the previous example, does not cause any probabilities to crop up. Either the sent 1 looks like a 1 to Eve or it does not. Similarly for sending no picture of a diagram of 1.

Acknowledgements: We thank Catherine Meadows for helpful discussions. Funding provided by ONR.

References

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

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

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

[4] Fred I. Dretsky. *Knowledge and the Flow of Information*. CSLI Publications, 1999.

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

[6] Ira S. Moskowitz, Richard E. Newman, Daniel P. Crepeau, and Allen R.Miller. Covert channels and anonymizing networks. In *Proceedings of WPES 2003*, pages 79–93. ACM Press, 2003.

[7] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.