The text that follows is an author's preprint of a paper published in *Journal of Experimental and Theoretical Artificial Intelligence*, December, 2018, DOI: 10.1080/0952813X.2018.1544282

A mosaic of Chu spaces and Channel Theory I: Category-theoretic concepts and tools

Chris Fields
23 Rue des Lavandières
11160 Caunes Minervois, FRANCE
fieldsres@gmail.com

and

James F. Glazebrook
Department of Mathematics and Computer Science
Eastern Illinois University, 600 Lincoln Ave.
Charleston, IL 61920–3099, USA
jfglazebrook@eiu.edu
Adjunct Faculty
Department of Mathematics
University of Illinois at Urbana–Champaign
Urbana, IL 61801, USA

November 28, 2018

Abstract

Chu Spaces and Channel Theory are well established areas of investigation in the general context of category theory when applied to semantically-based information flow. In this Part I of a two-part work, we review a range of related concepts and examples showing how these methods can be applied to logic and computer science, including Formal Concept Analysis, distributed systems and ontology development. We also discuss spatial coarse-graining in relationship to information, and in this direction we establish some basic simplicial and categorical techniques which will supplement the other methods of this Part I when they are applied to characterize visual object identification and the inference of mereological (i.e. part-whole) complexity in Part II.

Keywords: Chu space, Information Channel, Infomorphism, Formal Concept Analysis, Distributed System, Local Logic, Ontology, Cocone, Simplex, Colimit.

Contents

1 Introduction

Category theory has provided a language and a range of conceptual tools suited to the general study of complex process as they feature extensively in areas such as computer science, artificial intelligence, the life sciences, and the study of ontologies (reviewed by Baianu et al., 2006; Ehresmann and Vanbremeersch, 2007; Goguen, 2005a,b; Healy and Caudell, 2006; Healy, 2010; Poli, 2001, 2006; Porter, 1994; Rosen, 1986; Spivak and Kent, 2012). This follows a tradition in conceiving of a range of descriptive methods in philosophy as first advocated by F. Brentano (1981), and then later by E. Husserl (1970), and others (surveyed by e.g. Simons, 1987; Smith, 2003). The rigorous mathematical formalism of category theory has its origins in the work of Eilenberg and Mac Lane (1945); Mac Lane (1971) (see also, e.g. Awodey, 2010).

One particular categorical concept is that of a *Chu space*, which entered computer science as a representable model of linear logic originally formulated by Barr (1979, 1991) and Seely (1989). An advantage of using Chu spaces is their flexibility in adapting to a wide range of interpretations and applications. They are more general than topological spaces, and they can be represented in straightforward *object-attribute* rectangular/matrix-like arrays (the rows consisting of object names, and the columns consisting of attribute names; so an $[ij]$-entry simply means that an object o_i has an attribute a_j). From the observational perspective, the attributes are taken to provide information about the structural and dynamical configurations of and between objects (for example, see Hitzler, Hölldobler and Seda (2004)). Following earlier developments of the theory, Chu spaces emerged with importance in areas dealing with machine learning and data mining; these include (but are not limited to) parallel programming algorithms, information retrieval, concurrent computation automata, physical systems, local logics, formal concept analysis (see below), the semantics of observation-measurement problems, decision theory, and ontological engineering (Abramsky, 2012; Allwein, Yang and Harrison, 2011; Barr, 1979; Barwise and Seligman, 1997; Berners-Lee, Hendler and Lassila, 2001; Pratt, 1995, 1999a,b; Zhang and Shen, 2006). Accordingly, one may find a variety of interpretations of Chu-space representations, including the object/attribute criteria used to define informational relationships, depending on the chosen context. *

During information processing, the various channels of assimilation may possess intrinsic qualities that influence the type of inferences they derive, regarding the general premise that "X being A carries the information that Y is B" (Dretske, 1981). As a step towards conceptualizing information flow within a logical environment in category-theoretic terms, the basic elements of Chu spaces have been adapted to the concept of *Classifications*, as the latter are expressed in terms of *Tokens* and *Types* (Barwise and Seligman, 1997; Barwise, 1997, 1999; Barwise and Perry, 1983). The resulting framework of *Channel Theory* casts information flow within a logical and distributed

*For the reader's convenience, and to supplement the survey of Chu spaces from the point of view of theory and applications, we offer the link: http://chu.stanford.edu/guide.html systems environment. An *infomorphism*, as a reformulated Chu morphism (in a sense 'dual') constitutes a pivotal concept of Channel Theory, by defining a channel through which the information represented by one classification is re-represented in another.

We have two broad aims in this two-part work. In this Part I, we survey and assemble the main concepts and tools of Chu spaces and Channel Theory in a compact integrated form, and briefly review a scope of techniques and examples. In Part II (Fields and Glazebrook, 2018) we employ these tools and methods to develop an original category-theoretic description of human visual object identification, a cognitive process that has been deeply characterized by several decades of experimental and theoretical investigation and that is widely viewed as a model for higher cognition in general. We focus in particular on the construction of object files (Kahneman, Triesman and Gibbs, 1992) and object tokens (Zimmer and Ecker, 2010) as intermediate representations, the binding of type and token information in object categorization and identification (Martin, 2007; Fields, 2012; Keifer and Pulvermüller, 2012), and the recognition and categorization of mereologically-complex individuals. We show in Part II that the use of category-theoretic methods reveals deep dualities in the visual object identification process that suggest that it can be considered structurally and functionally scale-free, and advance the hypothesis that human cognition may be scale-free in general.

Here, we commence by defining and reviewing some of the basic properties of Chu spaces in §2. Although Chu spaces have been traditionally applied to fields such as those listed above, they also have a number of other significant applications of interest here. How Chu spaces can be implemented within Formal Concept Analysis and Domain Theory (e.g. to represent information systems and approximable concepts following Hitzler and Zhang, 2004; Krötzsch, Hitzler and Zhang, 2005; Scott, 1982; Zhang and Shen, 2006), is reviewed in §3. In §4 we discuss representations of spaces (and representations by spaces), spatial coarse-graining and finite sampling of information (Gratus and Porter, 2006; Sorkin, 1991a); we then review the representation of sampled information by simplicial complexes constructed "above" the sampled space in §5. The following two sections, §6 and §7, establish a similar working account of Channel Theory. We survey a number of motivating examples and applications, including Distributed Systems (Barwise and Seligman, 1997) in §7.2, the flow of information in Ontology Comparison and Alignment (Kalfoglou and Schorlemmer, 2004; Schorlemmer, 2002, 2005) in §7.3, Event Classifications §7.5, State Spaces and Cognizance Classifications (Sakahara and Sato, 2008, 2011) in §7.7. The category-theoretic concepts of *cocone* and *colimit* (e.g. Awodey, 2010) naturally arise in both Chu space and Channel Theory descriptions; we review these concepts in §8 with illustrative examples. This completes the assembly of the mathematical foundation to be applied in Part II. We expect that this Part I survey may in itself be of independent interest to researchers across the AI community.

2 Chu spaces and Chu transforms

2.1 Basic definitions for objects and attributes

Definition 2.1. *A (dyadic or two-valued) Chu space* C consists of a triple $(C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ where C_{o} is a set of *objects*, C_{a} is a set of *attributes*, along with a *satisfaction relation* (or *evaluation*) $\Vdash_{\mathsf{C}} \subseteq C_{\mathsf{o}} \times C_{\mathsf{a}}$.

For observational purposes, we may regard the "attributes" as providing information about the structural and dynamical configurations of and between the "objects." Two objects can be distinguished if, but only if, there is at least one attribute that they do not share. Otherwise, objects are said to be equivalent. This sense of equivalence formalizes Leibniz' principle of "identity of indiscernibles." The "objects" and "attributes" can equally well be thought of as "states" and "events," with "states" distinguished by the "events" that can occur in them or, as we will see, in terms of "tokens" and "types" or by other similar pairs of concepts.

Definition 2.2. A *morphism* or *Chu transform* of a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ to a Chu space $\mathsf{D} = (D_{\mathsf{o}}, \Vdash_{\mathsf{D}}, D_{\mathsf{a}})$ is a pair of functions $(f_{\mathsf{a}}, f_{\mathsf{o}})$ with $f_{\mathsf{o}} : C_{\mathsf{o}} \longrightarrow D_{\mathsf{o}}$, and $f_{\mathsf{a}} : D_{\mathsf{a}} \longrightarrow C_{\mathsf{a}}$, such that for any $x \in C_{\mathsf{o}}$, and $y \in D_{\mathsf{a}}$, we have $f_{\mathsf{o}}(x) \Vdash_{\mathsf{D}} y$, if and only if $x \Vdash_{\mathsf{C}} f_{\mathsf{a}}(y)$.

If $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ is a Chu space, then $\mathsf{C}^{\perp} = (C_{\mathsf{a}}, \Vdash_{\mathsf{C}}^{\mathrm{op}}, C_{\mathsf{o}})$ is the *dual space* of C in which the roles of objects and attributes are interchanged. This sense of duality allows us to think, for example, of attributes being distinguished by the objects to which they apply, events being distinguished by the states in which they participate, or types being distinguished by the tokens they include. Chu-space duality will provide, in Part II (Fields and Glazebrook, 2018), the key to representing recurrent networks in a fully-symmetric way.

Generally, for some set K, we have a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ over K, with a satisfaction relation (evaluation) $\Vdash_{\mathsf{C}}: C_{\mathsf{o}} \times C_{\mathsf{a}} \longrightarrow \mathsf{K}$, such that $\Vdash_{\mathsf{C}} (a, b)$ is an element of K. This leads to a convenient matrix representation of a Chu space with entries in K (see the basic examples below). Note that no structure on K is assumed. The category of Chu spaces over K along with their morphisms is habitually denoted as $\mathbf{Chu}(\mathbf{Set}, \mathsf{K})$.

2.2 Chu flows

What is the information preserved when switching between Chu spaces that are tied by a Chu transform? Let a *Chu flow* (van Benthem, 2000), cf. Barwise and Seligman (1997) be specified by a "flow formula" constructed from the elements of the following schema:

$$x \Vdash a \mid \neg(x \Vdash a) \mid \ \wedge \ \mid \vee \mid \exists x \mid \forall a. \tag{2.1}$$

Any such formula $\psi(a_1, \ldots, a_k, x_1, \ldots, x_m)$ specifies which objects x_i have which attributes a_i in the Chu space in which it applies. In van Benthem (2000) it was shown that for finite Chu spaces C and D, the existence of a Chu transform $\mathsf{C} \longrightarrow \mathsf{D}$ is equivalent to every flow formula valid in C being valid in D as well. The transform $\mathsf{C} \longrightarrow \mathsf{D}$ can, in this case, be viewed as "transporting" the information encoded in valid flow formulas from C to D; it can thus be thought of informally as a "channel" from C to D and as implicitly providing a sense of "spatial" and/or "temporal" separation between C and D. These informal notions will be made more precise in §6.

Example 2.1. A given flow formula $\psi(a_1, \ldots, a_k, x_1, \ldots, x_m)$ can give rise to useful relations between k attributes (*types*) and m objects (*tokens*). For instance on taking a_1, a_2 as types, and x_1, x_2 as tokens, we have (van Benthem, 2000):

$$\begin{aligned}
\forall x(\neg a_1 \in x \vee a_2 \in x) &\ \subset \ \text{object inclusion}, \\
\forall x(\neg a_1 \in x \vee \neg a_2 \in x) &\ \boxminus \ \text{object incompatibility}, \\
\exists a(a \in x_1 \wedge a \in x_2) &\ \mathsf{o}\ \text{type overlap}.
\end{aligned}$$

2.3 Biextensional collapse

Following Pratt (1999a) we define a pair of maps relative to power sets $\mathcal{P}(\cdot)$ as follows:

$$\begin{aligned} \hat{\alpha} : C_{\mathsf{o}} &\longrightarrow \mathcal{P}(C_{\mathsf{a}}) \text{ with } \hat{\alpha}(x) = \{a \in C_{\mathsf{a}} : x \Vdash_{\mathsf{C}} \mathsf{a}\} \\ \hat{\omega} : C_{\mathsf{a}} &\longrightarrow \mathcal{P}(C_{\mathsf{o}}) \text{ with } \hat{\omega}(\mathsf{a}) = \{x \in C_{\mathsf{o}} : x \Vdash_{\mathsf{C}} \mathsf{a}\}. \end{aligned} \tag{2.2}$$

Given $X \subseteq C_{\mathsf{o}}$, and $A \subseteq C_{\mathsf{a}}$, the above two maps extend to the following maps, respectively Zhang and Shen (2006):

$$\begin{aligned} \alpha : \mathcal{P}(C_{\mathsf{o}}) &\longrightarrow \mathcal{P}(C_{\mathsf{a}}) \text{ with } \alpha(X) = \{\mathsf{a} : \forall x \in X \ x \Vdash_{\mathsf{C}} \mathsf{a}\} \\ \omega : \mathcal{P}(C_{\mathsf{a}}) &\longrightarrow \mathcal{P}(C_{\mathsf{o}}) \text{ with } \omega(A) = \{x : \forall \mathsf{a} \in A \ x \Vdash_{\mathsf{C}} \mathsf{a}\}. \end{aligned} \tag{2.3}$$

A Chu space C is said to be *extensional* if $\hat{\omega}$ is injective, and *separable* if $\hat{\alpha}$ is injective. If C is both extensional and separable, then let us say it is *biextensional*. In fact, any Chu space can be turned into a biextensional type, provided the lack of injectivity of α and ω can be factored out in a suitable sense. This creates a *biextensional collapse* of a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$, namely the Chu space

$$\widehat{\mathsf{C}} = (\widehat{C}_{\mathsf{o}}, \Vdash_{\widehat{\mathsf{C}}}, \widehat{C}_{\mathsf{a}}) = (\hat{\alpha}(C_{\mathsf{o}}), \Vdash_{\widehat{\mathsf{C}}}, \hat{\omega}(C_{\mathsf{a}})), \tag{2.4}$$

where $\hat{\alpha}(x) \Vdash_{\widehat{\mathsf{C}}} \hat{\omega}(a)$, if and only if $x \Vdash_{\mathsf{C}} \mathsf{a}$.

Remark 2.1. Let us explain the above terms with regards to the matrix representation of a Chu space. 'Separable' means that all rows are distinct, and 'extensional' means that all columns are distinct. In the biextensional collapse, any repetitions in the rows of objects (tokens) and columns of attributes (types) are factored out, thus removing unnecessary repetitions in the content of information, and hence minimizing the amount of processing units in a given algorithm.

Example 2.2. Consider a topological space $(X, \mathcal{U})$, where X is the set of points, and $\mathcal{U}$ is the set of open sets. Taking $\Vdash = \in$, and $\mathsf{K} = \mathbf{2} = \{0, 1\}$, gives a Chu space $(X, \in, \mathcal{U})$. In the matrix representation, the space is extensional (no repeated columns). Relative to the set of open sets, the columns are closed under arbitrary union and finite intersection. Here the relationship is interpreted as: $x \Vdash U = 1$ implies $x \in U$, and $x \Vdash U = 0$ implies $x \notin U$. In this way the category **Top** of topological spaces along with their continuous maps, embeds, in a sense, $\mathbf{Top} \longrightarrow \mathbf{Chu}(\mathbf{Set}, \mathbf{2})$. For further specifics in the topological context, see Pratt (1999a).

Example 2.3. With $\mathsf{K} = \mathbf{2} = \{0, 1\}$, take a set (of objects) $X = \{a, b, c\}$. This can be represented as the Chu space

$\Vdash$								
a	0	1	0	1	0	1	0	1
b	0	0	1	1	0	0	1	1
c	0	0	0	0	1	1	1	1

We could take any 8-membered set A (of attributes) to index the columns. As pointed out in Pratt (1999a, Chap. 1), it is convenient to view the columns as self-identifying, with each column a function $X \to \mathsf{K}$, otherwise expressed as $A \subseteq \mathsf{K}^X$. Chu spaces so organized, with $\Vdash$ understood, are called *normal*. So we may just write (X, A) for $(X, \Vdash, A)$. For $\mathsf{K} = \mathbf{2}$ this is equivalent to viewing columns as subsets of X; otherwise said, the characteristic functions of those subsets, with 1's in the column as representing members of the subset, and the 0's non-members.

Suppose we delete three columns from above, so as to obtain

$\Vdash$					
a	0	1	1	1	1
b	0	0	1	0	1
c	0	0	0	1	1

If we define $a \leq b$ to be true just when this is the case in every column (as usual by taking $0 \leq 1$), then we now have $b \leq a$ and $c \leq a$, still three distinct members equipped with a nontrivial order relation. This relation endows the Chu space with the structure of a *partially ordered set*, or for short, a *poset*, denoted $(X, \leq)$. Specifically, for all $a, b, c \in X$, this means: i) $a \leq a$ (reflexivity); ii) $a \leq b \leq a$ implies $a = b$ (antisymmetry); iii) $a \leq b$ and $b \leq c$ implies $a \leq c$ (transitivity). Such a poset structure on a Chu space will later show up in §4.1 and §4.2.

Remark 2.2. In the following sections we will present some further examples and applications of Chu spaces. This will be especially the case in §6 and §7 when we outline the Channel Theory formulation of Chu spaces.

3 Formal Concept Analysis and Computation in Chu spaces

Category theory can be viewed as a unified language for handling conceptual complexities in both mathematics and computer science. Chu spaces and Chu flows provide a natural way of representing both the structure and processing of information and have been used to investigate the semantic foundations and design of data structures and programming languages. The examples that follow illustrate these applications and introduce concepts that will prove useful later.

3.1 Concept lattices and approximable concepts

Formal Concept Analysis (FCA) is an approach to the semantics of symbolic data structures that studies the clustering of attributes into partially ordered sets that give rise to a *concept lattice* (Ganter, Wille and Franzke, 1999). *Domain Theory* (DT) for programming languages is concerned with higher-order relations between concepts that involve partial information and successive approximation, and with the question of when information can be approximated by finitely representable *approximable concepts* (Zhang and Shen, 2006) (cf. *formal contexts* described in Hitzler and Zhang (2004)). A central idea of FCA is the distinction between the 'extension' of a concept as consisting of all objects belonging to that concept, and the 'intension' of the concept as consisting of all attributes common to all objects belonging to that concept. Defining a concept in FCA thus involves identifying a collection of attributes which agrees with the 'intension of the extension'. Note that the idea of 'intension' in FCA captures the philosophical notion of an "essential property" that all members (here, objects) of a category (here, a concept) must have.

This FCA notion of 'concept' has been shown to be intrinsic to a Chu space (Krötzsch, Hitzler and Zhang, 2005; Zhang and Shen, 2006); indeed each Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ has an associated complete lattice $\mathcal{L}\mathsf{C}$ of formal concepts associated with C. Zhang and Shen (2006, Th. 4.1) have further shown that for every complete lattice D of formal (in the sense of FCA) concepts, there is a Chu space C such that D is order-isomorphic to $\mathcal{L}\mathsf{C}$. The following definition(s) then characterize the differences between 'formal' (in the sense of FCA) and 'approximable' (in the sense of DT) concepts.

Let P, Q be sets, and $\mathcal{A} \subseteq \mathcal{P}(P)$, $\mathcal{B} \subseteq \mathcal{P}(Q)$ (recall that $\mathcal{P}(\cdot)$ denotes the power set). Any pair of functions $s : \mathcal{A} \longrightarrow \mathcal{B}$, $t : \mathcal{B} \longrightarrow \mathcal{A}$, is called a *Galois connection*, if for each $X \in \mathcal{A}$ and $Y \in \mathcal{B}$, $s(X) \supseteq Y$ if and only if $X \subseteq t(Y)$. With respect to a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$, we recall from (2.3) the two associated functions (depending on C, so $\alpha = \alpha_{\mathsf{C}}$ and $\omega = \omega_{\mathsf{C}}$):

$$\begin{aligned} \alpha &: \mathcal{P}(C_{\mathsf{o}}) \longrightarrow \mathcal{P}(C_{\mathsf{a}}) \text{ with } \alpha(X) = \{\mathsf{a} : \forall x \in X \ x \Vdash_{\mathsf{C}} \mathsf{a}\} \\ \omega &: \mathcal{P}(C_{\mathsf{a}}) \longrightarrow \mathcal{P}(C_{\mathsf{o}}) \text{ with } \omega(A) = \{x : \forall \mathsf{a} \in A \ x \Vdash_{\mathsf{C}} \mathsf{a}\}. \end{aligned} \tag{3.1}$$

The pair of maps (α, ω) forms such a Galois connection (Ganter, Wille and Franzke, 1999): i) the set of attribute (object) concepts of P forms a closure system, i.e. a family of subsets closed under intersection (Caspard and Monjardet, 2003); ii) the attribute (object) concepts of C under set inclusion form a complete lattice; and, iii) the lattice of attribute concepts, and the lattice of object concepts are anti-isomorphic to each other. We then have:

Definition 3.1.

(1) A subset $A \subseteq C_{\mathsf{a}}$ is called an *(formal) concept (of attributes)*, if it is a fixed point of $\alpha \circ \omega$, i.e. $\alpha(\omega(A)) = A$. Dually, a subset $X \subseteq C_{\mathsf{o}}$ is called a *(formal) concept (of objects)* if it is a fixed point of $\omega \circ \alpha$. For each object $x \in C_{\mathsf{o}}$, the set of its attributes $\alpha\{x\}$ is a concept.

(2) A subset $A \subseteq C_{\mathsf{a}}$ is called an *approximable concept*, if for every finite subset $X \subseteq A$, we have $\alpha(\omega(X)) \subseteq A$.

Note that (1) above allows "single-object concepts"; these will become important in Fields and Glazebrook (2018) as representations of object tokens (cf. (Fields, 2012)), as well as for a hierarchial iteration of the idea.

Definition 3.2. A *complete algebraic lattice* (henceforth, an *algebraic lattice*) is a partial order which is both a complete lattice and a directed complete partial order (dcpo).

We have now the following basic representation theorem for approximable concepts (Zhang and Shen, 2006, Th. 6.3):

Theorem 3.1. *For any Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$, the set of its approximable concepts $\mathcal{A}\mathsf{C}$ under inclusion forms an algebraic lattice. Conversely, for every algebraic lattice D, there is a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ such that D is order-isomorphic to $\mathcal{A}\mathsf{C}$.*

3.2 Chu spaces as information systems

An *information system* with "states" consisting of finite subsets of tokens selected from some set A can be defined in terms of an underlying Chu space as follows. Let $\mathsf{Fin}(A)$ be the set of finite subsets of A, and choose a subset $\mathsf{Con} \subset \mathsf{Fin}(A)$ and a relation $\vdash$ (see Scott, 1982, for details). Interpret the information states x (i.e. elements of Con) as objects, the tokens $a \in A$ as attributes, and let $x \Vdash a$ if and only if a is a member of x. In this case, the subset Con on A is called the *consistency predicate*, and $\vdash$ the *entailment relation*. Following Zhang and Shen (2006), a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ gives rise to an information system $(A_{\mathsf{C}}, \mathsf{Con}_{\mathsf{C}}, \vdash_{\mathsf{C}})$ via the assignment $A_{\mathsf{C}} = C_{\mathsf{a}}$, $X \vdash_{\mathsf{C}} \mathsf{a}$, if $\mathsf{a} \in \alpha_{\mathsf{C}} \circ \omega_{\mathsf{C}}(X)$, and a consistency predicate $\mathsf{Con}_{\mathsf{C}}$ for which every subset of C_{a} is consistent. Zhang and Shen (2006, Th. 4.6) have shown, for a given Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ with C_{a} finite, a state $X \subset C_{\mathsf{a}}$, taken to be a *concept*, is equivalent to X being a state of the derived information system $(A_{\mathsf{C}}, \mathsf{Con}_{\mathsf{C}}, \vdash_{\mathsf{C}})$. Intuitively, a Chu morphism in Definition 2.2 correlating the objects and attributes of C to those of some other Chu space D is a correlation between the respective information systems. Such a morphism similarly maps sequences of flow formulas valid in C to sequences of flow formulas valid in D, and hence correlates information *processes* in the respective information systems.

3.3 Ordered dynamical systems and computation

The stage is now set to develop a general notion of computation for arbitrary dynamical systems with discrete states. As a sequence of *measurements* of any arbitrary dynamical system can itself be considered a dynamical system with discrete states (Fields, 1989), nothing is lost by assuming discreteness. Artificial neural networks (ANNs) are such systems (Nauck, Klawonn and Kruse, 2003), as are Turing machines, cellular automata, etc.

Consider a quadruple $\langle S, \mathsf{ns}, \leq, \mathcal{T} \rangle$, where S is the state space of an information system as characterized above, ns is the next-state function, $\leq$ is a partial order and $\mathcal{T} : \mathcal{L} \longrightarrow S$ is a mapping where $\mathcal{L}$ denotes a propositional ('factual') language (Leitgeb, 2005). The map $\mathcal{T}$ assigns some proposition ϕ of $\mathcal{L}$ to each state $s \in S$; hence it represents the (stipulated) semantics of S. The action of ns, in this case, produces a sequence of propositions $\phi_0, \phi_1, \phi_2, ...$ and so can be interpreted as (in general, nonmonotonic) inference. If this sequence converges to some stable state ψ, the action of ns has "halted" and the proposition ψ can be interpreted as the "result" of the action of ns on ϕ_0. The design perspective in which ns is stipulated and the reverse engineering or debugging perspective in which ns must be discovered are both clearly supported within this picture.

Recasting the above in the language of K-valued Chu spaces provides a representation of computations with imprecise inputs, outputs or both. Recalling the attribute symbol $\Vdash$, let us define

$$\begin{cases} s \Vdash^t \phi & \text{iff } \mathcal{T}(\phi) = s \text{ (precise state information)}, \\ s \Vdash \phi & \text{iff } \mathcal{T}(\phi) \leq s \text{ (imprecise state information)}. \end{cases}$$

The computational interpretation is straightforward: $s \Vdash^t \phi$ if and only if ϕ completely specifies the system state, whereas $s \Vdash \phi$ if and only if the system state is described by ϕ as well as some other propositions in $\mathcal{L}$. A computation with an initial state $s \Vdash^t \phi$ and a final state $s' \Vdash \psi$, for example, would provide an ambiguous answer (ψ together with other propositions) to a precise question (ϕ).

This Chu/information space representation of computation has been adapted to capture Bayesian inference in a connectionist context (Dayan et al., 1995; McClelland, 1998); we develop this representation further in Part II (Fields and Glazebrook, 2018). The close relationship between Chu flows and infomorphisms as defined within Channel Theory (Barwise and Seligman, 1997) and their application to problems such as ontology alignment are considered in §6 and §7, respectively. In particular, state space systems will be further described in the context of Channel Theory in §7.5 and §7.7

4 Topology of information and observation

Propositions used to describe the world are semantically related; in the limit, all propositions in any language form a connected semantic network (Sowa, 2006). Observations or, more precisely, finitely-specifiable observational outcomes are similarly related. Considering an information system to be defined merely over a *set* of propositions provides no means of capturing such relations. It is, therefore, useful to introduce additional structure, with the addition of topological structure a natural first step. Doing this allows a structured notion of sampling the information encoded in a Chu space, and particularly the idea of a *finite sample of attributes* (FSA) for an object or collection of objects.

4.1 The Sorkin perspective

One approach to developing an information topology is via a notion of *causality*; this has been pursued by Sorkin (1991a,b) through the development of *causal sets*. While the motivation in this case has been to model the fundamental structure of spacetime in a way that could produce the continuum of macroscopic geometry as an emergent 'classical limit' (see Raptis and Zapatrin, 2001; Sorkin, 1991a,b, for details of the mathematical physics application domain), the techniques employed are generally applicable to approximating a class of highly structured or idealized spaces by means of taking a certain limit of less complex, more user-friendly spaces. The point is to represent non-spatial information in a spatial form as means of 'visualization', and to consider variation in data and observations in the context of such representations. The use of spatial dimensions as a way of "displaying" information in a meaningful way on both the input and output sides of connectionist systems (and more generally, ANNs) is an example of this approach. In this case, the very complex, essentially causal relations between information computed by a learning algorithm are approximated, on an imposed spatial array of output "units" that have no intrinsic spatial relationships, in a way that makes them meaningful to external observers (Rogers and McClelland, 2004, Ch. 8). An input array similarly approximates causal relations in "the world" when the array geometry is assigned semantic significance, e.g. in computer vision applications.

Fundamental causality relations between objects x, y, z can be expressed in terms of an order relation '$\prec$':

(1) $x \prec y \prec z \;\Rightarrow\; x \prec z$ (Transitivity: if y is the outcome of z, and x is the outcome of y, then x is the outcome of z).

(2) $x \prec y$ and $y \prec x \;\Rightarrow x = y$ (Antisymmetry).

(3) Let $[[x, y]]$ denote the cardinality of the number of elements z between x and y such that $x \prec z \prec y$, then $[[x, y]] < \infty$ (Discreteness).

These relations can also be expressed in terms of a (locally finite) poset as we do below; we then apply the inherent sense of causality to the structure provided by Chu spaces. This is achieved by approximating a highly structured space by a spatial model based on simplicial complexes and related posets as developed in Gratus and Porter (2006, 2005a,b), which we survey in part here, and in §5.

4.2 The Sorkin poset $P_{\mathcal{F}}$

Suppose we are given a topological space X, viewed as a space of 'observables', and let us observe X from a finite family of open sets (FFOS) $\mathcal{F}$, not necessarily covering X. This will represent *a set of observations made on* X, where objects are observed in relationship to their attributes. In this way, the FFOS partitions X into the 'attributes', and X can then be regarded as a union of 'zones' (see below) in which two points lie in the same zone if they share the very same attributes; in other words, they consist of clusters of points in the same open set of the FFOS, and thus cannot be distinguished by the corresponding set of observations.

We can define an equivalence relation "$\sim_{\mathcal{F}}$", by $x \sim_{\mathcal{F}} x'$, if and only if for all $U \in \mathcal{F},\ x \in U$ if and only if $x' \in U$. Thus two points are *equivalent* if all the observations from $\mathcal{F}$ attribute the same positive or negative result on both of them. This is simply another way of stating the causality relations above. We can factor out by this equivalence relation to obtain a quotient mapping:

$$\pi_{\mathcal{F}} : X \longrightarrow X_{\mathcal{F}} = X/ \sim_{\mathcal{F}} \tag{4.1}$$

where the quotient $X_{\mathcal{F}}$ can be regarded as encoding the observational data on X in a way that organizes that data by "merging" equivalent observations.

The space $X_{\mathcal{F}}$ has topological type T_0[†] and corresponds to a poset denoted $P_{\mathcal{F}}$ and constructed as follows. We take $[x]_{\mathcal{F}}$ to be the equivalence class of $x \in X$, with $[x]_{\mathcal{F}} \leq [y]_{\mathcal{F}}$ if and only if for every open set $U \in \mathcal{F}$, if $y \in U$, then $x \in U$. For practical purposes we consider the family $\mathcal{F}$ as finite, and $X_{\mathcal{F}}$ is a finite T_0-space. Each point $[x]_{\mathcal{F}}$ is contained in a minimal open set $U_{[x]}$ of $X_{\mathcal{F}}$, and $[x]_{\mathcal{F}} \leq [y]_{\mathcal{F}}$ if and only if $x \in U_{[y]}$. The resulting poset $P_{\mathcal{F}}$ contains much of the essential observational (or causal data) on X. Besides organizing that data, this poset will serve as a means of 'measurement' (though not point-dependent) for gauging whether 'objects' and 'attributes' (or, 'tokens' and 'types') are seen as proximate to each other, or in contrast, are actually very far apart. Its structure is motivated by the ideas of Sorkin (1991a), as adopted by Gratus and Porter (2006), describing how certain types of spaces can be approximated by 'inverse limits' of more regular spaces.

Observe that the FFOS $\mathcal{F}$ determines a secondary topology $\tau(\mathcal{F})$ on X which is just the topology generated by $\mathcal{F}$. If $\tau(X)$ denotes the original topology on X, then $\tau(\mathcal{F}) \subseteq \tau(X)$ with the closure with respect to $\tau(\mathcal{F})$ interpreted as a proximity between 'zones' (see below). Let $\tau(P_{\mathcal{F}})$ denote the quotient topology on $P_{\mathcal{F}}$ such that the map

$$\pi_{\mathcal{F}} : (X, \tau(\mathcal{F})) \longrightarrow (P_{\mathcal{F}}, \tau(P_{\mathcal{F}})), \tag{4.2}$$

is continuous (and then is seen to be an open map). We summarize the nomenclature in the following:

Definition 4.1. Given X and a FFOS $\mathcal{F}$, we say that the pair $(P_{\mathcal{F}}, \pi_{\mathcal{F}})$ is a *Sorkin model of X relative to $\mathcal{F}$*, in which case $P_{\mathcal{F}}$ is called the *Sorkin poset* for $(X, \mathcal{F})$. Given $x \in P_{\mathcal{F}}$, the corresponding subset $\pi_{\mathcal{F}}^{-1}(x) \subseteq X$ is called the *zone determined by x*, which in general will be neither an open nor closed subset of X.

Definition 4.2. Given two FFOSs $\mathcal{F}$ and $\mathcal{G}$ of a topological space X, we say that $\mathcal{F}$ is *a Sorkin refinement* of $\mathcal{G}$ if $\mathcal{G} \subseteq \tau(\mathcal{F})$.

[†] A topological space X is said to be a T_0-*space* if given distinct points of X, there is an open set of X that contains one but not the other. T_0-spaces naturally give rise to a partial order defined on the set of points of X, where $x \leq y$, if for each open set $U \subseteq X$, $y \in U$ implies $x \in U$, and conversely.

From Gratus and Porter (2006, Prop. 11) we observe that $\mathcal{F}$ is a refinement of $\mathcal{G}$, if and only if there exists a continuous surjective map $\pi_{\mathcal{F}\mathcal{G}} : P_{\mathcal{F}} \longrightarrow P_{\mathcal{G}}$ such that the following diagram commutes

$$\begin{tikzcd}
(X, \tau(X)) \arrow[r, "\pi_{\mathcal{F}}"] \arrow[dr, "\pi_{\mathcal{G}}"'] & (P_{\mathcal{F}}, \tau(P_{\mathcal{F}})) \arrow[d, "\pi_{\mathcal{F}\mathcal{G}}"] \\
& (P_{\mathcal{G}}, \tau(P_{\mathcal{G}}))
\end{tikzcd} \tag{4.3}$$

that is, $\pi_{\mathcal{G}} = \pi_{\mathcal{F}\mathcal{G}} \circ \pi_{\mathcal{F}}$.

4.3 A Chu FSA

Firstly, we recall the elementary result that any space X along with a FFOS $\mathcal{F}$ can be formulated in terms of a Chu space $\mathsf{C} = (X, \in, \mathcal{F})$, such that an object $x \in X$ satisfies an attribute $U \in \mathcal{F}$, if $x \in U$ (see Example 2.2); hence C is a normal Chu space (Pratt, 1999a). Thinking back to §2.3, we see that the quotient map $\pi_{\mathcal{F}}$ in (4.2) is simply the universal map to the biextensional collapse of $\mathsf{C} = (X, \in, \mathcal{F})$, and that the Chu space $\mathsf{C}_{P_{\mathcal{F}}} = (\mathcal{F}, \in, \tau(P_{\mathcal{F}}))$ is itself biextensional, observing that the poset structure on $P_{\mathcal{F}}$ is given by

$$\hat{\alpha}(x) \leq \hat{\alpha}(y) \Longleftrightarrow \forall a \in C_{\mathsf{a}} \; (y \Vdash_{\mathsf{C}} a \Rightarrow x \Vdash_{\mathsf{C}} a) \Longleftrightarrow \hat{\alpha}(x) \supseteq \hat{\alpha}(y). \tag{4.4}$$

To avoid possible complications, we assume, as in (Gratus and Porter, 2006), that $\mathcal{F}$ is suitably 'sampled' and extensional (that is, $\mathcal{F}$ has no repetitive columns). Accordingly, we obtain a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ consisting of a finite sample of attributes $\mathcal{F}$ resulting in a pair $(\mathsf{C}, \mathcal{F})$, entitled a *Chu FSA*. Given $(\mathsf{C}, \mathcal{F})$, we call $\mathsf{C}_{|\mathcal{F}} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, \mathcal{F})$ the *corestriction* of $(\mathsf{C}, \mathcal{F})$.

4.4 Putting a topology on a Chu space

The next step is to put a topology on a Chu space C. Thus, we commence by saying that C is *topologically closed* if the attributes C_{a} is a topology on objects C_{o}, meaning that C is normal, and C_{a} includes all unions and finite intersections. Without too much loss of generality, we assume that C is biextensional. Thus given C, we have a topologically closed Chu space

$$\tau(\mathsf{C}) = (C_{\mathsf{o}}, \in, \tau(C_{\mathsf{a}})), \tag{4.5}$$

which is naturally a topological closure of C. Furthermore, there is a universal Chu morphism $\tau : \tau(\mathsf{C}) \longrightarrow \mathsf{C}$, with $\tau_{\mathsf{o}} : C_{\mathsf{o}} \longrightarrow C_{\mathsf{o}}$ the identity, and $\tau_{\mathsf{a}} : C_{\mathsf{a}} \longrightarrow \tau(C_{\mathsf{a}})$ the inclusion. The point here is that $\tau(\mathsf{C})$ contains the same informational (or observational) structure as the original C, and in $\tau(\mathsf{C})$ the information has been encoded by means of the propositional operations of geometric logic, and sampled via $\mathcal{F} \subset C_{\mathsf{a}}$. Hence, as proposed by Gratus and Porter (2006, 2005b), *a Sorkin model $\mathsf{C}_{\mathcal{F}}$ for $(\mathsf{C}, \mathcal{F})$ is defined to be the biextensional collapse\Sorkin poset of $\mathsf{C}_{|\mathcal{F}}$*. Gratus and Porter (2006, Prop. 18) have also shown that any row x in $\mathsf{C}_{\mathcal{F}}$ consists of n entries 0 or 1, and thus corresponds to a flow formula (§2.2):

$$(x \Vdash a_{i_1}) \wedge \cdots \wedge (x \Vdash a_{i_k}) \wedge \neg(x \Vdash a_{i_{k+1}}) \wedge \cdots \wedge \neg(x \Vdash a_{i_n}), \tag{4.6}$$

in turn showing that the rows of $\mathsf{C}_{\mathcal{F}}$ can be considered to encode the elementary flow formulae:

$$\exists x(\bigwedge_{i \in \mathcal{F}_1} (x \Vdash a_i) \wedge \bigwedge_{i \in \mathcal{F}_2} \neg(x \Vdash a_i)), \tag{4.7}$$

for given partitions $(\mathcal{F}_1, \mathcal{F}_2)$ of $\mathcal{F}$.

Given Chu spaces $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$ and $\mathsf{D} = (D_{\mathsf{o}}, \Vdash_{\mathsf{D}}, D_{\mathsf{a}})$, we say that C is a *Sorkin refinement of* D if there exists a Chu transform $\phi : \tau(\mathsf{C}) \longrightarrow \mathsf{D}$, which is the identity on objects ($\phi_{\mathsf{o}}(x) = x$). Further, any Chu space is a Sorkin refinement of itself, the Sorkin refinement is transitive, and if C is both extensional and a Sorkin refinement of D, then the map ϕ is uniquely determined (Gratus and Porter, 2006, Prop. 20).

5 Introducing simplicial methods on Chu spaces

Given a topology on an information space, algebraic methods can be used to extend the topology into a geometry. The resulting (discrete) geometry provides a natural representation of sets of observations made at different resolutions or scales, and hence a natural way to represent coarser- to finer-grained approximations of the topology. Such approximations will provide, in Fields and Glazebrook (2018, §4), the basis for a mereotopology of "parts" of objects.

5.1 Simplicial complexes: basic definitions

We first introduce simplicial complexes as representations of observational data, following Cordier and Porter (1989); Friedman (2012); Goerss and Jardine (1999) and Spanier (1966).

Definition 5.1. A *simplicial complex* K consists of a set K_0 of objects called the *vertices* and a set of finite, non-empty subsets of K_0 called the *simplices*. The latter satisfy the condition that if $\sigma \subset K_0$ is a simplex, and if $\tau \subset \sigma$ (with $\tau \neq \emptyset$), then τ is also a simplex. Simplicial complexes are objects in a category denoted **Simpl**.

Simplicial complexes over an information space provide the structure needed to define an information geometry. To each simplicial complex K is associated the *polyhedron* or *geometric realization* of K, denoted $|K|$, formed from the set of all functions $K_0 \longrightarrow [0, 1]$ satisfying:

i) if $\alpha \in |K|$, then the set $\{v \in K_0 : \alpha(v) \neq 0\}$ is a simplex of K;

ii) $\sum_{v \in K_0} \alpha(v) = 1$.

Here each function α can be thought of as "picking out" a subset of vertices to be the vertices of some particular polyhedron. These functions are normalized so that they "pick out" each vertex to the same extent.

For any simplex $s \in K$, there is an associated set $|s| = \{\alpha \in |K| : \alpha(v) \neq 0 \Rightarrow v \in s\}$ as well as a set $\langle s \rangle = \{\alpha \in |K| : \alpha(v) \neq 0 \Leftrightarrow v \in s\}$. Often $\alpha(v)$ is called the v^{th} *barycentric coordinate of* α, and the mapping $|K| \longrightarrow [0, 1]$ defined by $p_v(\alpha) = \alpha(v)$ is the v^{th} *barycentric projection of* α. With these coordinates, a metric d can be defined on K as given by

$$d(\alpha, \beta) = \Big(\sum_{v \in K_0} (p_v(\alpha) - p_v(\beta))^2 \Big)^{\frac{1}{2}}. \tag{5.1}$$

This distance $d(\alpha, \beta)$ measures the number of vertices shared between the polyhedra "picked out" by α and β, normalized to account for differences in the numbers of vertices of the two polyhedra.

Definition 5.2. If K and L are two simplicial complexes, a simplicial mapping $f : K \longrightarrow L$ is a map $f_0 : K_0 \longrightarrow L_0$ of vertex sets that preserves simplices, meaning that if $\sigma \subset K_0$ is a simplex of K, then its image $f(\sigma) \subset L_0$ is a simplex of L.

Any simplicial complex K gives rise, in a straightforward way, to a poset, namely the poset of its "faces." The elements of this poset are the simplices of K, arranged according to the rule $\sigma \leq \rho$ if σ is a face of ρ, that is, if $\sigma \subseteq \rho$ as subsets of the vertex set K_0. Note that unions of "adjoining" faces are faces with this definition. We will tacitly employ the (contravariant) functor relating the categories **Simpl** $\longrightarrow$ **Sets** to speak of a *simplicial set* corresponding to its underlying structure as a simplicial complex.

5.2 Simplicial homotopy

Two simplicial maps $f, g : X \longrightarrow Y$ of simplicial sets X, Y are said to be *homotopic* if there exists a simplicial map $H : X \times I \longrightarrow Y$ (here $I = [0, 1]$ the closed unit interval) such that $H_{|X\times\{0\}} = g$, and $H_{|X\times\{1\}} = f$. In other words, we have $g = H \circ i_0$, and $f = H \circ i_1$, with respect to inclusion maps $i_0 : X \times \{0\} \hookrightarrow X \times I$, and $i_1 : X \times \{1\} \hookrightarrow X \times I$. This is summarized by the following commutative diagram

$$\begin{tikzcd} X \times \{1\} \arrow[d, "i_1"'] \arrow[dr, "f"] & \\ X \times I \arrow[r, "H"] & Y \\ X \times \{0\} \arrow[u, "i_0"] \arrow[ur, "g"'] & \end{tikzcd} \tag{5.2}$$

so that we have $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$.[‡]

5.3 The nerve of a relation

For a space X and open cover $\mathcal{F}$ of X, the *Čech nerve* $N(\mathcal{F})$ *of* $\mathcal{F}$ is defined as the simplicial complex whose vertices are the (open) sets in $\mathcal{F}$ and for which $\{U_0, \ldots, U_n\}$ is an n-simplex of $N(\mathcal{F})$ if and only if $\bigcap_{i=0}^{n} U_i \neq \emptyset$. Intuitively, the Čech nerve is the simplicial complex over $\mathcal{F}$ comprising only *connected* simplices. As pointed out by Gratus and Porter (2006), the face poset $P_{\mathcal{F}}$ as defined above bears a close relation with $N(\mathcal{F})$, but they need not be identified. In a dual sense, there is the *Vietoris complex* $V(\mathcal{F})$ *of* $(X, \mathcal{F})$ in which the vertices are simply the points of X, and $\langle x_0, \ldots, x_n\rangle$ is an n-simplex if there exists a $U \in \mathcal{F}$ that contains them all, that is, $\{x_0, \ldots, x_n\} \subseteq U$.

Dowker (1952) provides an abstraction in this setting, given a *relation* $\mathcal{R} \subseteq X \times Y$ from X to Y. A simplicial complex $K_{\mathcal{R}}$, called *the nerve of the relation* can be specified by: i) the vertices of $K_{\mathcal{R}}$ are those elements $x \in X$ for which there exists a y such that $(x, y) \in \mathcal{R}$, and ii) the set $\{x_0, \ldots, x_n\} \in X$ is an n-simplex if and only if there exists some y such that $(x_i, y) \in \mathcal{R}$, for $0 \leq i \leq n$. From this it can be deduced that $N(\mathcal{F})$ and $V(\mathcal{F})$ each provide the same information about the open cover $\mathcal{F}$ up to homotopy.

[‡]We adopt this natural definition of a simplicial homotopy as found in Friedman (2012); Goerss and Jardine (1999). In Friedman (2012) it is compared with the traditional, more technically oriented definition as seen in other textbooks on the subject.

Example 5.1. Let us exemplify some of these concepts for the basic case of the circle S^1, following Porter (2002). Here we take an open covering $\mathcal{F} = \{U_1, U_2, U_3\}$, relative to polar coordinates, with

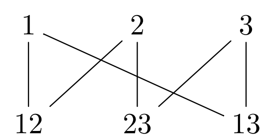
$$\begin{aligned}
U_1 &= \big(-\frac{2\pi}{3}, \frac{2\pi}{3}\big);\\
U_2 &= \text{“}\big(0, -\frac{2\pi}{3}\big)\text{” i.e. } \big(0, \pi\big] \cup \big(-\pi, -\frac{2\pi}{3}\big);\\
U_3 &= \text{“}\big(\frac{2\pi}{3}, 0\big)\text{” i.e. } \big(\frac{2\pi}{3}, \pi\big] \cup \big(-\pi, 0\big).
\end{aligned} \tag{5.3}$$

Every point of S^1, with the exception of 0, $\frac{2\pi}{3}$ and $-\frac{2\pi}{3}$, is in exactly two of these, with a total of six equivalence classes. Choosing three representatives for the non-singleton classes gives the following minimal open sets:

$$U_0 = U_1,\ U_{\frac{2\pi}{3}} = U_2,\ U_{-\frac{2\pi}{3}} = U_3$$

$$U_{\frac{\pi}{3}} = U_1 \cap U_2 := U_{12},\ U_{-\frac{\pi}{3}} = U_1 \cap U_3 := U_{13},\ U_{\pi} = U_2 \cap U_3 := U_{23}.$$

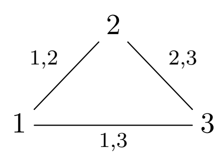
Now we have a partially ordered set with associated Hasse diagram



(5.4)

showing that $S^1_{\mathcal{F}}$ has 6 points, and the homotopy type of the former is that of S^1.

If we set $X = S^1$ and take the open cover $\mathcal{F} = \{U_1, U_2, U_3\}$ as above, the vertices of $N(\mathcal{F}) = \langle U_1 \rangle, \langle U_2 \rangle, \langle U_3 \rangle$, and the 1-simplices of $N(\mathcal{F}) = \langle U_1, U_2 \rangle, \langle U_1, U_3 \rangle, \langle U_2, U_3 \rangle$. Thus, $N(\mathcal{F})$ may be represented schematically by the diagram



(5.5)

Recalling that any simplicial complex determines a poset by subset inclusion of simplices, it can be seen that the resulting poset is the opposite of that representing $X_{\mathcal{F}}$.

5.4 The Čech and Vietoris nerves of a Chu space

In the context of a Chu space $\mathsf{C} = (C_{\mathsf{o}}, \Vdash_{\mathsf{C}}, C_{\mathsf{a}})$, the *Čech nerve* is the simplicial complex denoted $N(\mathsf{C})$ with vertex set C_{a} and where a (non-empty) subset $\{a_0, \ldots a_p\}$ of C_{a} is a p-simplex if there is an object $x \in C_{\mathsf{o}}$ satisfying $x \Vdash_{\mathsf{C}} a_i$, for $0 \leq i \leq p$. This is motivated by the fundamental principle that for simplicial complexes, *the nerve can be viewed as a set of instructions serving to construct (an approximation of) a space by fitting together the individual geometric simplicies.*[§] At the same time, the associated *Vietoris nerve* $V(\mathsf{C})$ is, in this context, the Čech nerve of the dual space $\mathsf{C}^{\perp}$. Given that some $\{a_0, \ldots a_p\}$ comprises a simplex, the latter can be symbolized as $\langle a_0, \ldots a_p \rangle$.

[§]The nerve specifies, in effect, which simplices adjoin each other by “sharing an edge.” Porter (2002) provides a number of illustrative examples providing an intuition leading to the discussion in Gratus and Porter (2006).

As pointed out by Gratus and Porter (2006, §3) there may be possible complications in dealing with induced Chu morphisms, since the set C_{a} may be infinitely large. To remedy this situation, it is necessary to finitely sample the attributes by restricting consideration to a subset $\mathcal{F}$ of C_{a}. Thus following Gratus and Porter (2006, Prop. 4), if $f = (f_{\mathsf{o}}, f_{\mathsf{a}}) : \mathcal{P} = (P_{\mathsf{o}}, \Vdash_{\mathcal{P}}, P_{\mathsf{a}}) \longrightarrow \mathcal{Q} = (Q_{\mathsf{o}}, \Vdash_{\mathcal{Q}}, Q_{\mathsf{a}})$ is a morphism of Chu spaces, and $\mathcal{F}$ is an open cover (the *observations*) representing a finite sample of the attributes of $\mathcal{Q}$, i.e we have $\mathcal{F} \subseteq Q_{\mathsf{a}}$ and finite, then there is an induced map

$$f = (f_{\mathsf{o}}, f_{\mathsf{a}}) : \mathcal{P} = (P_{\mathsf{o}}, \Vdash_{\mathcal{P}}, f_{\mathsf{a}}(\mathcal{F})) \longrightarrow \mathcal{Q} = (Q_{\mathsf{o}}, \Vdash_{\mathcal{Q}}, \mathcal{F}). \tag{5.6}$$

Furthermore, there are also induced simplicial maps given as follows: firstly, with respect to the Vietoris nerve, we have

$$V(f) : V(P_{\mathsf{o}}, \Vdash_{\mathcal{P}}, f_{\mathsf{a}}(\mathcal{F})) \longrightarrow V(Q_{\mathsf{o}}, \Vdash_{\mathcal{Q}}, \mathcal{F}), \tag{5.7}$$

given by $V(f)\langle p_0, \ldots, p_n \rangle = \langle f_{\mathsf{o}}(p_0), \ldots, f_{\mathsf{o}}(p_n) \rangle$, and secondly, for any choice of splitting, the function $f_{\mathsf{a}} : \mathcal{F} \longrightarrow f_{\mathsf{a}}(\mathcal{F})$ (recall $\mathcal{F} \subseteq Q_{\mathsf{a}}$) induces a simplicial map with respect to the Čech nerve

$$N(f) : N(P_{\mathsf{o}}, \Vdash_{\mathcal{P}}, f_{\mathsf{a}}(\mathcal{F})) \longrightarrow N(Q_{\mathsf{o}}, \Vdash_{\mathcal{Q}}, f_{\mathsf{a}}(\mathcal{F})), \tag{5.8}$$

given by $N(f)\langle f_{\mathsf{a}}(q_0), \ldots, f_{\mathsf{a}}(q_n) \rangle = \langle q_0, \ldots, q_n \rangle$.

5.5 The Chu FSA and induced morphisms between nerves

The above simplicial procedures show that, for any Chu FSA $(\mathsf{C}, \mathcal{F})$, there are two associated simplicial complexes $N(\mathsf{C}_{|\mathcal{F}})$ and $V(\mathsf{C}_{|\mathcal{F}})$, along with the associated posets of their faces. Recall that we took $\mathsf{C}_{\mathcal{F}}$ to denote the biextensional collapse\Sorkin poset of $\mathsf{C}_{|\mathcal{F}}$. Let $\widehat{\mathcal{F}}$ denote a corresponding family of attributes. Again assuming $\mathsf{C}_{|\mathcal{F}}$ is extensional (no repeated columns in $\mathcal{F}$) then (Gratus and Porter, 2006, Th. 21), the quotient map

$$\pi_{\mathcal{F}} : \mathsf{C}_{|\mathcal{F}} \longrightarrow \mathsf{C}_{\mathcal{F}}, \tag{5.9}$$

exists, and there is an induced isomorphism

$$\pi^N_{\mathcal{F}} : N(\mathsf{C}, \mathcal{F}) \xrightarrow{\cong} N(\mathsf{C}_{|\mathcal{F}}, \widehat{\mathcal{F}}), \tag{5.10}$$

of simplicial complexes.

Intuitively, representing a set of observations of an FSA of a Chu space by a simplicial complex renders it a coarse-graining of the underlying Chu space, with the "grain size" determined by the number of vertices in the Čech nerve. The adjacency relations implicit in the Čech nerve provide this coarse-graining with a local geometry. This construction thus captures the important intuitions that 1) observations in practice always have finite resolution, and hence yield finitely-specifiable outcomes, 2) any coarse-graining can be further coarse-grained by incorporating subsets of mutually-connected simplices at the finer scale into single simplices at the new coarser scale, and 3) any coarse-graining can be finitely refined by reversing this process. The hierarchy of coarse-grainings produced becomes, in Fields and Glazebrook (2018), a mereological hierarchy describing a complex object constructed out of parts that have both adjacency relations and local geometric relations.

6 An excursion into Channel Theory I

6.1 Classifications: Tokens and Types

Situation Theory and Channel Theory (Barwise and Seligman, 1997) provide a structure for describing informational relations in the setting of information flow through systems distributed across space and time. They provide a conceptual and schematic generalization of the ontological notion of information as a causal connection introduced by Dretske (1981), i.e. of the intuitive notion that a state of some system B can *encode* or *carry* information about some system A to which it is coupled by a physical interaction that serves as a channel. An assumption lending realism to the approach is that the channels through which information flows may have implicit or unknown properties that alter in some way the information flowing through them. When such alterations are systematic, they can be considered *inferences* from input information to output information that are implemented by the channel. It is important to note that channels conceived of in this way are interposed between observations (i.e. systems viewed as recording observational outcomes); hence the inferences implemented by a channel must be *inferred* by comparing the output with the input. Channels can thus be identified with elementary gates in either classical or quantum (Nielsen and Chaung, 2000) computations.

The fundamental concept of Channel Theory is the idea of a classification relating tokens to the types that encompass them.

Definition 6.1. A *classification* $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ consists of a set $\mathrm{Tok}(\mathcal{A})$ consisting of the *tokens of* $\mathcal{A}$, a set $\mathrm{Typ}(\mathcal{A})$ consisting of the *types of* $\mathcal{A}$, and a classification relation

$$\Vdash_{\mathcal{A}} \subseteq \mathrm{Tok}(\mathcal{A}) \times \mathrm{Typ}(\mathcal{A}), \tag{6.1}$$

that classifies tokens to types.

A classification $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ has the structure of a Chu space, that is, via the assignment $(\mathrm{Tok}(\mathcal{A}), \Vdash_{\mathcal{A}}, \mathrm{Typ}(\mathcal{A})) \mapsto (\text{object}, \Vdash, \text{attribute})$ or, as is more typical in Barwise and Seligman (1997), the 'dual' form $(\mathrm{Tok}(\mathcal{A}), \Vdash_{\mathcal{A}}, \mathrm{Typ}(\mathcal{A})) \mapsto (\text{attribute}, \Vdash, \text{object})$. As will be seen in §6.5 below, these interpretations are interchangeable. Let us also keep in mind that for Chu spaces, "objects" and "attributes" can be aptly replaced by terms such as "events" and "states", with $\Vdash$ then interpreted as selecting the events that occur in a given state or, alternatively, the states participating in a given event.

Remark 6.1. As in §2.1, we can take a classification $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ over a set K, with evaluation $\Vdash_{\mathcal{A}} \subseteq \mathrm{Tok}(\mathcal{A}) \times \mathrm{Typ}(\mathcal{A}) \longrightarrow \mathsf{K}$, where $\Vdash_{\mathcal{A}} (a, b)$ is an element of K.

Instances of Chu spaces (such as Example 2.2) conveniently carry over to classifications, and conversely. In the following, we present some examples that were originally formulated within Channel Theory.

Example 6.1. Following Allwein, Moskowitz and Chang (2004), let

$$\mathbf{FOL} = \langle Models, Sentences, \Vdash_{\mathbf{FOL}} \rangle,$$

where *Sentences* are sentences in First Order Logic (FOL). $Models$ are models of FOL sentences, and $x \Vdash_{\mathbf{FOL}} S$, if and only if x is a model of the sentence S. Here, there are various internal relations holding on both the set of sentences and that of models, but none are imposed as external conditions in this case without further modification. One could also reverse matters, by taking the Types to be Models, and the Tokens as Sentences, so that Sentences in this case would be classified by Models.

6.2 Infomorphisms

Here we recall the idea of a Chu morphism in order to link the information between two given classifications $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ and $\mathcal{B} = \langle \mathrm{Tok}(\mathcal{B}), \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}} \rangle$. In this case it is useful to define "switching relations" $\overrightarrow{f} : \mathrm{Typ}(\mathcal{A}) \longrightarrow \mathrm{Typ}(\mathcal{B})$ and $\overleftarrow{f} : \mathrm{Tok}(\mathcal{B}) \longrightarrow \mathrm{Tok}(\mathcal{A})$ that can be specified by introducing the Channel Theory concept of an *infomorphism*. Specifically:

Definition 6.2. Given two classifications $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ and $\mathcal{B} = \langle \mathrm{Tok}(\mathcal{B}), \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}} \rangle$, an *infomorphism* $f : \mathcal{A} \rightleftarrows \mathcal{B}$, is a pair of contravariant maps

i) $\overrightarrow{f} : \mathrm{Typ}(\mathcal{A}) \longrightarrow \mathrm{Typ}(\mathcal{B})$

ii) $\overleftarrow{f} : \mathrm{Tok}(\mathcal{B}) \longrightarrow \mathrm{Tok}(\mathcal{A})$

such that for all $b \in \mathrm{Tok}(\mathcal{B})$, and for all $a \in \mathrm{Typ}(\mathcal{A})$, we have

$$\overleftarrow{f}(b) \Vdash_{\mathcal{A}} a, \text{ if and only if } b \Vdash_{\mathcal{B}} \overrightarrow{f}(a). \tag{6.2}$$

This last condition may be schematically represented by:

$$\begin{array}{ccc} \mathrm{Typ}(\mathcal{A}) & \xrightarrow{\overrightarrow{f}} & \mathrm{Typ}(\mathcal{B}) \\ {\scriptstyle \Vdash_{\mathcal{A}}} \Big| & & \Big| {\scriptstyle \Vdash_{\mathcal{B}}} \\ \mathrm{Tok}(\mathcal{A}) & \xleftarrow{\overleftarrow{f}} & \mathrm{Tok}(\mathcal{B}) \end{array} \tag{6.3}$$

Note that this definition, given in Barwise and Seligman (1997), employs the 'dual' interpretation of types as objects and tokens as attributes. Interpreting tokens as objects and types as attributes yields infomorphisms with the usual Chu-morphism arrow directions.

Remark 6.2. In the context of situations, 'attributes' can be interpreted as statements of 'situation types'. In the Dretske spirit, to say that "x is T_1" transmits information that "y is T_2" can be represented as an infomorphism representing these classification statements. Here the content of information such as (T_1, T_2) is defined as the 'type', and the carrier of the respective types, such as (x, y), is defined as the 'token'.

Example 6.2. Let $\mathbf{M} = \langle Messages, Contents, \Vdash_{\mathbf{M}} \rangle$ where Messages are classified by their Contents (Allwein, Moskowitz and Chang, 2004). Suppose we have another such classification $\mathbf{M'} = \langle Messages', Contents', \Vdash_{\mathbf{M'}} \rangle$. An infomorphism $f : \mathbf{M} \longrightarrow \mathbf{M'}$ may represent a function decoding messages from $\mathbf{M'}$ to messages in $\mathbf{M}$, so that whatever can be noted about the translation, may be mapped into something noted in the original message. That is, $m^f \Vdash_{\mathbf{M}} C \Leftrightarrow m \Vdash_{\mathbf{M'}} C^f$.

Example 6.3. Here is an example from Decision Theory (Allwein, Yang and Harrison, 2011, §2.3). Let $\mathbf{S}$ be a classification of propositional logic and its model states, and let f represent a decision which evaluates a state s and the agent making the decision (e.g. "either walk home, or take the bus home"). Let $\mathbf{O}$ be the classification of outcomes, and let s^f represent a particular outcome of the decision of choosing either option (either "walk home" or "take a bus home"). A proposition, denoted Q over outcomes (in accordance with a slogan such as "Keeping Fit") characterizes them. Let Q^f be the proposition categorizing all of the states in which Q is satisfied. Thus, with respect to the above scheme of infomorphisms, we set the classification $\mathcal{A} = \mathbf{O}$, and $\mathcal{B} = \mathbf{S}$, and (6.3) thus leads to

$$\begin{tikzcd} \mathrm{Typ}(\mathbf{O}) \arrow[r, "\overrightarrow{f}"] \arrow[d, dash, "\Vdash_{\mathbf{O}}"'] & \mathrm{Typ}(\mathbf{S}) \arrow[d, dash, "\Vdash_{\mathbf{S}}"] \\ \mathrm{Tok}(\mathbf{O}) & \mathrm{Tok}(\mathbf{S}) \arrow[l, "\overleftarrow{f}"'] \end{tikzcd} \tag{6.4}$$

in which case the infomorphism condition is expressed by $s \Vdash_{\mathbf{S}} Q^f$ if and only if $s^f \Vdash_{\mathbf{O}} Q$.

Remark 6.3. It is worth noting that in the framework of infomorphisms, there is a natural mapping between tokens A and the set of informational states:

$$A \longrightarrow \mathsf{S}(A). \tag{6.5}$$

For instance, the truth classification of a first order language L is the classification whose types are the sentences of L, and the tokens are the L-structures. In which case, the classification relation is defined by $N \Vdash \varphi$, if and only if φ is true in the structure of L (see (Barwise and Seligman, 1997, Example 4.6)).

6.3 Information channels

An *information channel* **Chan** consists of an indexed family $\{f_i : \mathcal{A}_i \rightleftarrows \mathbf{C}\}_{i \in \mathcal{I}}$ of infomorphisms having a common codomain $\mathbf{C}$ called the *core of the channel* **Chan**:

$$\begin{tikzcd} & \mathbf{C} & & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[u, "f_2"'] & \ldots\ \mathcal{A}_i \arrow[ul, "f_i"'] \ \ldots \end{tikzcd} \tag{6.6}$$

The core $\mathbf{C}$ is essentially a carrier of information flow between the f_i and hence between the classifications $\mathcal{A}_i$, and is itself a classification in the above sense. The tokens $\mathrm{Tok}(\mathbf{C})$ of $\mathbf{C}$ are called *connections*. A connection c is said to *connect* the tokens $f_i(c)$ of the classifications $\mathcal{A}_i$ for $i \in \mathcal{I}$ (note that tokens are mapped from $\mathbf{C}$ to the $\mathcal{A}_i$ in the 'dual' interpretation of Barwise and Seligman (1997)). A channel with index set $\{0, \ldots, n-1\}$ is called an *n-ary* channel. Composing information channels amounts to taking their limit and the channels themselves may be refinable by straightforward categorical means (Barwise and Seligman, 1997).

The above definition extends the intuitive picture of a channel as a wire connecting two agents (i.e. classifiers) to the idea of a blackboard, or other shared memory, via which multiple classifiers exchange information. The shared memory $\mathbf{C}$ being itself a classifier provides it with a structure that can affect how information is written to, and read from it; one can imagine, for example, a "smart" blackboard that incorporates a function such as multi-language translation. Analogous conceptualizations will be treated schematically in the descriptive mechanism presented in Part II. Consistent with the essentially causal notion of information of Dretske (1981), the connections between the tokens of different classifiers are purely functional; no overarching semantics is assumed. How such a semantics can be constructed, *post hoc*, given a channel is discussed in §7.3 below.

6.4 Cocone of infomorphisms

A network of infomorphisms between classifications admits a limit classification that gathers all of the information in the network into a single classification (a *cone*) with projections back down to the individual classifications (Barwise and Seligman, 1997). There is a dual notion which we will describe as follows. A channel is an instance of the more general category-theoretic concept of a *cocone* being the core classification. To motivate the construction, consider any finite directed graph with vertex labels $1, 2, ..., n$ and edge labels f_{ij}. Considering such a graph to represent a network of communicating agents is, from a category-theoretic perspective, invoking a map G (technically, a functor from the category of finite directed graphs to the category of classifications) that constructs a classification $G(i)$ at each vertex and an infomorphism $G(f_{ij})$ at each edge. A *commuting finite cocone* of infomorphisms (e.g. Barwise and Seligman (1997); Allwein, Yang and Harrison (2011)) is a finite network of classifications $G(i)$ and infomorphisms $G(f_{ij})$, a vertex classification $\mathbf{C}$, and a collection of infomorphisms $g_i : G(i) \longrightarrow \mathbf{C}$:

$$\begin{tikzcd}
 & & \mathbf{C} & & \\
G(1)\ldots \arrow[urr, "g_1"] & G(i) \arrow[ur, "g_i"'] \arrow[rr, "G(f_{ij})"'] & & G(j) \arrow[ul, "g_j"] & \ldots G(n) \arrow[ull, "g_n"']
\end{tikzcd} \tag{6.7}$$

The commutativity condition is that for all f_{ij}, we have $g_i = g_j \circ G(f_{ij})$. The base of the cocone consists of the classifications and infomorphisms constructed by G; the *cocone vertex classification* $\mathbf{C}$ together with the maps g_i, is a channel. Note that in the complementary sense, a *commuting finite cone* of infomorphisms consists of a finite network of classifications $G(i)$ and infomorphisms $G(f_{ji})$, a vertex classification $\mathbf{C}$, and a collection of infomorphisms $g_i : G(i) \longrightarrow \mathbf{C}$. For all f_{ji}, we have $g_i = G(f_{ji}) \circ g_j$, and all arrows in the above diagram are reversed.

In short, we have this *colimit classification* into which there are infomorphisms from each constituent classification, and this colimit contains all of the information that is common to the different component parts of the network. The generalization from channel to cocone will prove useful in the discussion of "minimal covers" of distributed systems in §7.2. We further characterize cocones and relate them to colimits in the descriptive discussion of §8. In Part II of this work (Fields and Glazebrook (2018)) we will apply these concepts to model both abstraction-based and mereological categorization as well as the process of tracking individual category members through time as both their features and their contexts of observation change.

6.5 The flip of a classification

For any classification $\mathcal{A}$, the *flip* of $\mathcal{A}$, is the classification $\mathcal{A}^{\perp}$ whose tokens are the types of $\mathcal{A}$, whose types are the tokens of $\mathcal{A}$, such that $\alpha \Vdash_{\mathcal{A}^{\perp}} a$ if and only if $a \Vdash_{\mathcal{A}} \alpha$ (see Barwise and Seligman (1997, §4.4)). In deciding how to model a classification there may be epistemological questions, e.g.

the types in question are given as things or attributes we may know about, and the tokens are those things we wish to have information about. The fact that the flip of a classification is a classification (and both can be treated as Chu spaces) and these behave well under infomorphisms, means that a situation involving types or tokens, can be dualized to tokens or types. For instance, "the type set of a token" dualizes to "the token set of a type". Effectively, $f : \mathcal{A} \rightleftarrows \mathcal{B}$ is an informorphism if and only if $f^{\perp} : \mathcal{B}^{\perp} \rightleftarrows \mathcal{A}^{\perp}$ is an infomorphism (Barwise and Seligman, 1997, Prop. 4.19). Further, $(\mathcal{A}^{\perp})^{\perp} = \mathcal{A}$ with $(f^{\perp})^{\perp} = f$, and $(fg)^{\perp} = g^{\perp} f^{\perp}$ (Barwise and Seligman, 1997, Prop. 4.20). Thus for $f, f^{\perp}$ respectively, we have the commuting diagrams

$$\begin{array}{ccc} \mathrm{Typ}(\mathcal{A}) & \xrightarrow{\overrightarrow{f}} & \mathrm{Typ}(\mathcal{B}) \\ \Vdash_{\mathcal{A}} \Big| & & \Big| \Vdash_{\mathcal{B}} \\ \mathrm{Tok}(\mathcal{A}) & \xleftarrow{\overleftarrow{f}} & \mathrm{Tok}(\mathcal{B}) \end{array} \qquad \begin{array}{ccc} \mathrm{Tok}(\mathcal{B}) & \xrightarrow{\overleftarrow{f}} & \mathrm{Tok}(\mathcal{A}) \\ \Vdash_{\mathcal{B}}^{-1} \Big| & & \Big| \Vdash_{\mathcal{A}}^{-1} \\ \mathrm{Typ}(\mathcal{B}) & \xleftarrow{\overrightarrow{f}} & \mathrm{Typ}(\mathcal{A}) \end{array} \tag{6.8}$$

6.6 The nerve of a classification

As any classification is a Chu space, any operation defined for Chu spaces is meaningful for a classification. A finite sample $\mathcal{F}$ of 'attributes', for example, becomes a finite sample of 'tokens' (or 'types'). Simplicial complexes are defined as in §5.1 and nerves as in §5.4. The Čech nerve of a classification $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$, for example, is the simplicial complex $N(\mathcal{A})$ with vertex set $\mathrm{Tok}(\mathcal{A})$, where a (non-empty) subset $\{b_0, \dots b_p\}$ of $\mathrm{Tok}(\mathcal{A})$ is a p-simplex if there is a type $v \in \mathrm{Typ}(\mathcal{A})$ satisfying $v \Vdash_{\mathcal{A}} b_i$, for $0 \leq i \leq p$. The notion of the Vietoris nerve follows in a similar way as in §5.4.

If $\mathcal{F} = \mathrm{Tok}(\mathcal{B})$ is a finite sample of tokens of a classification $\mathcal{B}$, we have the infomorphism (6.3):

$$f = (\overleftarrow{f}, \overrightarrow{f}) : (\overleftarrow{f}(\mathcal{F}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}}) \longrightarrow (\mathcal{F}, \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}}), \tag{6.9}$$

while for a finite sample $\mathcal{G} \subseteq \mathrm{Typ}(\mathcal{A})$, we have:

$$f = (\overleftarrow{f}, \overrightarrow{f}) : (\mathrm{Tok}(\mathcal{A}), \mathcal{G}, \Vdash_{\mathcal{A}}) \longrightarrow (\mathrm{Tok}(\mathcal{B}), \overrightarrow{f}(\mathcal{G}), \Vdash_{\mathcal{B}}), \tag{6.10}$$

We will henceforth assume that finite samples of tokens (and types) have been taken, so that we may consider, as in (5.8), well-defined simplicial maps

$$N(f) : N(\mathcal{A}) \longrightarrow N(\mathcal{B}), \tag{6.11}$$

as defined for the Čech nerve of the corresponding Chu spaces, here with respect to finite samples of tokens and types. Here again, the driving intuition is that no more than a finite number of tokens are ever available *to observation* as exemplars of any type, and no more than a finite number of properties (i.e. types) are ever available *to observation* as characteristics of token. The notions of coarse-graining and finite refinement introduced in §5.5 carry over to this setting exactly.

6.7 Associating a theory with a classification

Here, and in the following sections, we collect together some useful definitions from Barwise and Seligman (1997) and Barwise (1997), starting with *sequents* and *theories*:

Definition 6.3. Let Σ be an arbitrary set (which may be viewed as set of *types*). A binary relation $\vdash$ between subsets of Σ is called a *consequence relation on* Σ. A *(Gentzen) sequent* is a pair $I = \langle \Gamma, \Delta \rangle$ of subsets of Σ (here it is apt to view Γ and Δ as sets of *situation types*). A sequent $I = \langle \Gamma, \Delta \rangle$ is said to hold of a situation s provided that if s supports every type in Γ, then it supports some type in Δ. A sequent I is said to be *information* about a set S of situations if it holds at each $s \in S$; here again the causal notion of information flow is evident. Finally, a sequent is called a *partition* of a set Σ' if $\Gamma \cup \Delta = \Sigma'$ and $\Gamma \cap \Delta = \emptyset$.

Definition 6.4. A *theory* is a pair $T = \langle \Sigma, \vdash_T \rangle$, where $\vdash_T$ is a consequence relation on Σ. A *constraint* of the theory T is a sequent $\langle \Gamma, \Delta \rangle$ of Σ for which $\Gamma \vdash_T \Delta$. A sequent $\langle \Gamma, \Delta \rangle$ is T-*consistent* if $\Gamma \nvdash_T \Delta$.

Here again, the idea of some aspects of a situation, either causally requiring or merely causally allowing other aspects of a situation, makes this definition clear.

Each classification has a theory associated with it in the following way (see also Definition 6.6 below). A *theory* $\mathsf{Th}(\mathcal{A}) = (\Sigma_{\mathcal{A}}, \vdash_{\mathcal{A}})$ generated by a classification $\mathcal{A}$, satisfies for all types α and all sets $\Gamma, \Gamma', \Delta, \Delta', \Sigma', \Sigma_0, \Sigma_1$ of types (Barwise and Seligman, 1997, Prop 9.5):

(1) *Identity*: $\alpha \vdash \alpha$.

(2) *Weakening*: If $\Gamma \vdash \Delta$, then $\Gamma, \Gamma' \vdash \Delta, \Delta'$.

(3) *Global cut*: If $\Gamma, \Sigma_0 \vdash \Delta, \Sigma_1$, for each partition $\langle \Sigma_0, \Sigma_1 \rangle$ of Σ, then $\Gamma \vdash \Delta$.

More generally, we can say that a theory $T = \langle \Sigma, \vdash_T \rangle$ is *regular* if it satisfies the above three conditions.

6.8 Local logics

We can specify a classification of a regular theory T as given by:

(1) $\mathrm{Typ}(\mathsf{Cl}(T)) = \mathrm{Typ}(T)$.

(2) $\mathrm{Tok}(\mathsf{Cl}(T)) = \{\langle \mathit{\Gamma}, \mathit{\Delta} \rangle : \langle \mathit{\Gamma}, \mathit{\Delta} \rangle$ is a T consistent partition of $\mathrm{Typ}(T)\}$.

(3) $\langle \Gamma, \Delta \rangle \vdash_{\mathsf{Cl}(T)} \alpha$ if and only if $\alpha \in \Gamma$.

Indeed, for any regular theory it can be seen that $\mathsf{Th}(\mathsf{Cl}(T)) = T$.

Since we are mainly considering distributed systems, and information processing entails computation within a logical framework, the following system of local logics (Barwise and Seligman, 1997, Def. 12.1) is one suited to representing various types of state spaces.

Definition 6.5. *A local logic* consists of a triple $(\mathcal{L} = \langle \mathrm{Tok}(\mathcal{L}), \mathrm{Typ}(\mathcal{L}), \Vdash_{\mathcal{L}} \rangle, \vdash_{\mathcal{L}}, \mathsf{N}_{\mathcal{L}})$ in which we have:

(1) a classification $\mathcal{L} = \langle \mathrm{Tok}(\mathcal{L}), \mathrm{Typ}(\mathcal{L}), \Vdash_{\mathcal{L}} \rangle$,

(2) a regular theory $\mathrm{Th}(\mathcal{L}) = (\mathrm{Typ}(\mathcal{L}), \vdash_{\mathcal{L}})$, and

(3) a subset $\mathsf{N}_{\mathcal{L}} \subset \mathrm{Tok}(\mathcal{L})$, called *the normal tokens* of $\mathcal{L}$, which satisfy all of the constraints of the theory $\mathrm{Th}(\mathcal{L})$ in (2).

Definition 6.6. Let $\mathcal{A}$ be a classification. The *local logic generated by* $\mathcal{A}$, denoted $\mathsf{Lg}(\mathcal{A})$, has classification $\mathcal{A}$, a regular theory $\mathsf{Th}(\mathcal{A}) = (\mathrm{Typ}(\mathcal{A}), \vdash_{\mathcal{A}})$, and all its tokens are normal. A logic is said to be *natural* if it is generated by some classification.

In fact, for any local logic $\mathcal{L}$ on $\mathcal{A}$, we have $\mathcal{L} = \mathsf{Lg}(\mathcal{A})$ by Barwise and Seligman (1997, Prop. 12.7). This relationship will be exemplified in Example 7.1 in the context of ontologies.

Intuitively, a local logic is "local" to the classification that generates it. Infomorphisms allow mapping the local logic of one classification to that of another; hence we can think of channels as supporting the flow of locally-defined logical relations between classifications. Recalling from §5.5 that any classification can be interpreted as defining a coarse-graining and hence a "scale" at which information is being organized and represented, each local logic can be thought of as a "logic at some level of description." In the context of cognition, this interpretation will be made explicit in Part II (Fields and Glazebrook, 2018) as applying to "levels" of either abstraction-based or mereological hierarchies.

As any classification can also be interpreted as describing a state space (§3.2), one can further associate a canonical logic $\mathsf{Lg}(\mathsf{S})$ to any state space S. Specifically, if S is such a state space with a classification of events $\mathsf{Evt}(\mathsf{S})$, then we can speak of an S-logic as a logic $\mathfrak{L}$ on this classification such that $\mathsf{Lg}(\mathsf{S}) \subseteq \mathfrak{L}$, with the intuition that this S-logic can accommodate the theory that is implicit to the structure of S (Barwise and Seligman, 1997, §16).

Remark 6.4. In Barwise (1997), $\mathcal{L}$ is called an *information context* and $\vdash$ is a binary relation relating sets of situation types. In this case $\mathsf{N}_{\mathcal{L}}$ is said to be a set of *normal situations*. Intuitively, these are the situations that the available information concerns. They may comprise all, or only some of the situations satisfying the information. For instance, we may start with some set of normal situations accounting for an individual's experiences to date, and then the information context consists of all the sequents satisfied by, i.e. consistent with, this experience. Stepping outside of the context generates "surprise" in the sense of expectation violation (cf. Friston (2010)).

Next, we look to what extent an infomorphism between classifications will respect the associated local logics. This is given by the following (Barwise and Seligman (1997, 12.3)):

Definition 6.7. A *logic infomorphism* $f : \mathcal{L}_1 \leftrightarrows \mathcal{L}_2$, consists of a covariant pair $f = \langle f^\frown, f^\vee \rangle$ of functions satisfying

(1) $f : \mathsf{Cl}(\mathcal{L}_1) \leftrightarrows \mathsf{Cl}(\mathcal{L}_2)$ is an infomorphism of classifications.

(2) $f^\frown : \mathsf{Th}(\mathcal{L}_1) \longrightarrow \mathsf{Th}(\mathcal{L}_2)$ is a theory interpretation, and

(3) $f^\vee[\mathsf{N}_{\mathcal{L}_2}] \longrightarrow \mathsf{N}_{\mathcal{L}_1}$

For further consequences of this definition, see Barwise and Seligman (1997).

Example 6.4. For instance, if we have a binary channel $\mathbf{Chan} = \{f : \mathcal{A} \rightleftarrows \mathbf{C}, g : \mathcal{B} \rightleftarrows \mathbf{C}\}$, then the local logic (see §6.8 below) on $\mathcal{B}$ induced by $\mathbf{Chan}$, is the logic $\mathsf{Lg}_{\mathbf{Chan}}(\mathcal{B}) = g^{-1}[f[\mathsf{Lg}(\mathcal{A})]]$ (in Barwise and Seligman (1997, 14.1) classifications $\mathcal{A}$ and $\mathcal{B}$ are interpreted as "idealization" and "reality", respectively). This induced logic can be characterized by (Barwise and Seligman, 1997, Prop. 14.2):

(i) A partition $\langle \Gamma, \Delta \rangle$ of $\mathrm{Typ}(\mathcal{B})$ is consistent in $\mathsf{Lg}_{\mathbf{Chan}}(\mathcal{B})$ if and only if $\langle f^{-1}[g[\Gamma]], f^{-1}[g[\Delta]] \rangle$ is the state description of some $a \in \mathrm{Tok}(\mathcal{A})$.

(ii) A token $b \in \mathrm{Tok}(\mathcal{B})$ is normal in $\mathsf{Lg}_{\mathbf{Chan}}(\mathcal{B})$, if and only if it is connected to some token $a \in \mathrm{Tok}(\mathcal{A})$.

6.9 Boolean Classification

We can exemplify local logics following Barwise and Seligman (1997); Barwise (1999) in terms of a Boolean classification $\mathcal{A} = \langle S, \Sigma, \Vdash, \wedge, \neg \rangle$. Here we have a set S of *situations* (tokens as objects) and a set Σ of *propositions* (types as attributes). This leads to a *Boolean local logic* $\mathcal{L} = \langle \mathcal{A}, \vdash, N \rangle$ where $N \subseteq S$ consists of the normal situations (Barwise and Seligman, 1997; Barwise, 1999). If $s \in N$ is a normal situation, $\Gamma \vdash \Delta$ and $s \Vdash p$, for all $p \in \Gamma$, then $s \Vdash q$ for some $q \in \Delta$. A partial ordering "$\subseteq$" on local logics $\mathcal{L}_1, \mathcal{L}_2$ on a fixed classification of $\mathcal{A}$ is defined by $\mathcal{L}_1 \subseteq \mathcal{L}_2$, if and only if:

(1) for all sets Γ, Δ of propositions, $\Gamma \vdash_{\mathcal{L}_1} \Delta$ entails $\Gamma \vdash_{\mathcal{L}_2} \Delta$, and

(2) every situation of $\mathcal{A}$ that is normal in $\mathcal{L}_2$ is also normal in $\mathcal{L}_1$.

Now, if we take any set of sequents T, the logic $\mathsf{Lg}(\mathcal{A}^T)$ generated by T on $\mathcal{A}$:

(i) has as normal situations all of those situations that satisfy the sequents in T,

(ii) has as constraints all sequents satisfied by all situations in N, and

(iii) has as normal situations all of the situations of $\mathcal{A}$ satisfying these constraints.

Given a fixed Boolean classification (Barwise, 1999):

(1) If $T_0 \subseteq T_1$, then $\mathsf{Lg}(\mathcal{A}^{T_0}) \subseteq \mathsf{Lg}(\mathcal{A}^{T_1})$.

(2) If $N_0 \supseteq N_1$, then $\mathsf{Lg}(\mathcal{A}_{N_0}) \subseteq \mathsf{Lg}(\mathcal{A}_{N_1})$.

Example 6.5. Suppose $\mathcal{A}$ is a classification of bird sightings (observations), and N consists of the actual sightings to date. Then $\mathsf{Lg}(\mathcal{A}_N)$ has as constraints all sequents satisfied by all those bird sightings to date, and the normal situations consist of all bird sightings that satisfy all of these constraints, a set that clearly contains N. This logic may entail the constraint BIRD $\vdash$ FLY, a constraint that holds as long as the situations encountered are meaningfully compatible with the elements of N. But now, suppose a penguin is observed. It will lie outside of the normal situations since it violates BIRD $\vdash$ FLY. This observation uncovers a new set $N' \supset N$, and accordingly their logics satisfy $\mathsf{Lg}(\mathcal{A}_{N'}) \subseteq \mathsf{Lg}(\mathcal{A}_N)$. There will be fewer constraints tenable in $\mathsf{Lg}(\mathcal{A}_{N'})$ as we can see, since BIRD $\nvdash$ FLY in this new logic.

7 An excursion into Channel Theory II

Classifications and channels have been applied widely in theoretical computer science; we briefly review some of these applications here as motivations for applying these tools to perceptual processing.

7.1 The information channel in a MLP network

One of the earliest ANNs studied was the *Multilayer Perceptron* (MLP) network (Rosenblatt, 1961; Rumelhart et al., 1986). As with typical ANNs, it has an input layers (I_i), hidden layers (H_i), and an output layer (O_i) with weighted directional (in an MLP, exclusively feedforward) linkages between subsequent layers. Kikuchi et al. (2003) develop a Chu space/Channel Theory representation of a 3-layer MLP, showing how the synaptic weights between layers form a channel; we follow their example closely.

Let $\mathbf{wo}_{ij}$ denote a synaptic weight between the j-th neuron in the hidden layer and the i-th neuron in the output layer. Similarly, let $\mathbf{wh}_{jk}$ denote a synaptic weight between the k-th neuron in the input layer and the j-th neuron in the hidden layer. Then for a given state function $f(x)$, the layers O_i, H_i and I_i are related in accordance with

$$O_i = f(\sum_j \mathbf{wo}_{ij} H_j) = f(\sum_j \mathbf{wo}_{ij}\ f(\sum_k \mathbf{wh}_{jk} I_k)). \tag{7.1}$$

It is convenient to regard an MLP with a fixed topology as a map $\mathbf{F} : \mathbf{I} \longrightarrow \mathbf{O}$, from the input data space $\mathbf{I} = \{I_i\}$ to the output data space $\mathbf{O} = \{O_i\}$, so that $\mathbf{F}$ is uniquely defined by a point in the parameter space of weights $\mathbf{\Phi} = \{\langle \mathbf{wh}, \mathbf{wo} \rangle\}$. In this way, a fixed topology on a MLP can be represented as $\mathbf{F}_{\langle \mathbf{wh}, \mathbf{wo} \rangle}$, once given $\langle \mathbf{wh}, \mathbf{wo} \rangle \in \mathbf{\Phi}$.

Next consider the sub-parameter spaces $\mathbf{\Phi}_h = \{\langle \mathbf{wh} \rangle\}$ and $\mathbf{\Phi}_o = \{\langle \mathbf{wo} \rangle\}$, and the following three classifications:

(1) $\mathcal{A} = (\mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}})$ (the states of "cognition" i.e. of $\mathbf{O}$);

(2) $\mathcal{B} = (\mathrm{Tok}(\mathcal{B}), \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}})$ (the states of the "environment" i.e. of $\mathbf{I}$);

(3) $\mathcal{C} = (\mathrm{Tok}(\mathcal{C}), \mathrm{Typ}(\mathcal{C}), \Vdash_{\mathcal{C}})$ (the states of the network);

where for the tokens $A = \mathbf{\Phi}_h,\ B = \mathbf{\Phi}_o$ and $C = \mathbf{\Phi}$, we define projections

$$\begin{aligned} g_h &: \mathbf{\Phi} \longrightarrow \mathbf{\Phi}_h,\ \langle \mathbf{wh}, \mathbf{wo} \rangle \mapsto \langle \mathbf{wh} \rangle \\ g_o &: \mathbf{\Phi} \longrightarrow \mathbf{\Phi}_o,\ \langle \mathbf{wh}, \mathbf{wo} \rangle \mapsto \langle \mathbf{wo} \rangle \end{aligned} \tag{7.2}$$

as well as the obvious respective inclusions $f_h : \mathbf{\Phi}_h \longrightarrow \mathbf{\Phi},\ f_o : \mathbf{\Phi}_o \longrightarrow \mathbf{\Phi}$. Thus, we obtain a core (and vertex of a cocone) that is in $\mathcal{C}$ along with an information channel

$$\begin{tikzcd} & \mathcal{C} = \langle \mathbf{wh}, \mathbf{wo} \rangle & \\ \langle \mathbf{wh} \rangle \arrow[ur, "f_h"] & & \langle \mathbf{wo} \rangle \arrow[ul, "f_o"'] \end{tikzcd} \tag{7.3}$$

where, as shown by Kikuchi et al. (2003), an algorithm for modifying $\langle \mathbf{wh}, \mathbf{wo} \rangle$ corresponds to a local logic on $\mathcal{C}$. This method can be developed in terms of Distributed Systems, as explained below in §7.2 (see also Remark 7.1).

7.2 Distributed Systems

Following the development of Barwise and Seligman (1997, Ch. 6) we provide an example of Dretske's "Xerox principle", namely, that information flow is transitive. Consider two information channels sharing common infomorphisms. Suppose the first channel represents the examination of a map, capturing the notion of a person's perceptual state carrying information about the map being examined. The second channel represents the informational relationship between the map and the region it depicts. These coupled channels can be illustrated:

$$\begin{tikzcd} & \mathbf{B}_1 & & \mathbf{B}_2 & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & & \mathcal{A}_2 \arrow[ul, "f_2"'] \arrow[ur, "f_3"] & & \mathcal{A}_3 \arrow[ul, "f_4"'] \end{tikzcd} \tag{7.4}$$

Recall that the elements of $\mathrm{Tok}(\mathbf{B}_1)$ are 'connections'. In this case the connections are spatio-temporal perceptual events involving persons in $\mathrm{Tok}(\mathcal{A}_1)$ looking at maps in (i.e. elements of) $\mathrm{Tok}(\mathcal{A}_2)$. The connections of the second channel, elements of $\mathrm{Tok}(\mathbf{B}_2)$, are spatio-temporal events involving making the maps in $\mathrm{Tok}(\mathcal{A}_2)$ to represent various regions in $\mathrm{Tok}(\mathcal{A}_3)$. Under certain circumstances, a person's perceptual state carries information about a particular mountain, given that the person is reading a map showing that mountain. In this regard, we may reasonably consider $\mathcal{A}_1$ as the *idealized space* of the physical space A_3.

The next step is to construct another channel that fits both $\mathbf{B}_1$ and $\mathbf{B}_2$ together. The process is: i) choose a person, ii) go to a map she is reading, and then iii) proceed to the region shown on that map. Here we will restrict to pairs $c = (b_1, b_2)$ ((perceptual event, map-making)), so that $f_2(b_1) = f_3(b_2) = a_2$ holds, i.e. there is just one map in question. In this way, types $\beta_1 = f_2(\alpha_2)$ and $\beta_2 = f_3(\alpha_2)$) are equivalent since they are both translations of a_2. This is built into the channel by identifying β_1 and β_2 (cf. the biextensional collapse of a Chu space) and gives rise to a new classification $\mathcal{C}$ having the above tokens and $\mathrm{Typ}(\mathcal{C}) = \mathrm{Typ}(\mathbf{B}_1 \cup \mathbf{B}_2)$, but identifying types originating from a common type in $\mathcal{A}_2$. Thus we obtain a new channel with core $\mathbf{C}$, connecting $\mathbf{B}_1$ and $\mathbf{B}_2$, as depicted below:

$$\begin{tikzcd} & & \mathbf{C} & & \\ & \mathbf{B}_1 \arrow[ur, "g_1"] & & \mathbf{B}_2 \arrow[ul, "g_2"'] & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & & \mathcal{A}_2 \arrow[ul, "f_2"'] \arrow[ur, "f_3"] & & \mathcal{A}_3 \arrow[ul, "f_4"'] \end{tikzcd} \tag{7.5}$$

The channel infomorphisms are defined via composition $h_1 = g_1 f_1$, and $h_3 = g_2 h_4$, so linking $\mathcal{A}_1$ to $\mathcal{A}_3$.

In general, we have

Definition 7.1. A *Distributed System* $\mathcal{A}$ consists of an indexed family $\mathsf{Cl}(\mathcal{A}) = \{\mathcal{A}_i\}_{i \in \mathcal{I}}$ of classifications, together with a set $\mathsf{Inf}(\mathcal{A})$ of infomorphisms having both domain and codomain in $\mathsf{Cl}(\mathcal{A})$. Each classification may be taken to support a local logic, along with the core of the channel.

Definition 7.2. An information channel **Chan** $= \{h_i : \mathcal{A}_i \rightleftarrows \mathbf{C}\}$ *covers* a distributed system $\mathcal{A}$ if $\mathsf{Cl}(\mathcal{A}) = \{\mathcal{A}_i\}_{i \in \mathcal{I}}$, and for each $i, j \in \mathcal{I}$, and for each infomorphism $f : \mathcal{A}_i \rightleftarrows \mathcal{A}_j$ in $\mathsf{Inf}(\mathcal{A})$, the following diagram commutes

$$\begin{tikzcd} & \mathbf{C} & \\ \mathcal{A}_i \arrow[ur, "h_i"] \arrow[rr, "f"'] & & \mathcal{A}_j \arrow[ul, "h_j"'] \end{tikzcd} \tag{7.6}$$

Chan is said to be a *minimal cover* if it covers $\mathcal{A}$, and for every other channel $\mathcal{D}$ covering $\mathcal{A}$ there is a unique infomorphism from $\mathbf{C}$ to $\mathbf{D}$.

A minimal cover of a system $\mathcal{A}$ converts the entire distributed system, consisting of multiple channels, into a single channel. Every distributed system has a minimal cover, and this cover is unique up to isomorphism (Barwise and Seligman, 1997, Th. 6.5).

Remark 7.1. The constructions of §7.1 and §7.2 can be extended to parallel distributed processing (PDP) and to more general multi-layer, bidirectional ANNs (see e.g. Dawson, 2005; Rogers and McClelland, 2004; Rumelhart et al., 1986). They are also applicable for modelling constructs of the massively parallel, competitively based, distributed system of the *Global Neuronal Workspace* (GNW) as studied in Baars and Franklin (2003); Dehaene and Naccache (2001); Dehaene and Changeux (2004); Wallace (2005). Such modelling can be implemented, for example, by the *Learned Intelligent Distribution Agent* (LIDA) architecture (Baars and Franklin, 2003; Franklin and Patterson, 2006; Friedlander and Franklin, 2008). As observed by Maia and Cleeremans (2005), feedforward and feedback projections in connectionist networks can engage selective attention toward more salient inputs, producing yet stronger weighting, that can predict which of the competing elements will gain access to the GNW central core (cf. Friston, 2010; Grossberg, 2013; Shanahan, 2012). We employ these methods to characterize bidirectional information and constraint flow between visual object files, object tokens, and event files in Fields and Glazebrook (2018).

7.3 Ontologies

It is often useful, when describing events or processes in some domain, to represent the ontology of the domain explicitly as a type hierarchy. Following Schorlemmer (2005) (cf. Kalfoglou and Schorlemmer, 2003, 2004):

Definition 7.3. An *ontology* is a tuple $\mathcal{O} = (C, \leq, \perp, |)$ where

(1) C is a finite set of concept symbols;

(2) $\leq$ is a reflexive, transitive, and anti-symmetric relation on C (a partial order);

(3) $\perp$ is a symmetric and irreflexive relation on C (disjointness);

(4) $|$ is symmetric relation on C (coverage).

Remark 7.2. This is a basic working definition by Schorlemmer (2005). In the case of *reference ontologies*, Kalfoglou and Schorlemmer (2003) append this definition with i) a finite set R of relations, and ii) a function $\sigma : R \longrightarrow C^+$ assigning to each relation its *arity*. This corresponds to the functor $(-)^+$ which sends a set C to the set of finite tuples whose elements are in C (see the example below).

In applications, the concepts in C typically characterize concrete objects in the domain, which are brought into the theory by populating $\mathcal{O}$ with tokens. Let X be a set of objects to be classified in terms of the concept symbols in C, via a classification relation $\Vdash_{\mathcal{A}}$; we define a classification $\mathcal{A} = \langle X, C, \Vdash_{\mathcal{A}} \rangle$, where $X = \mathrm{Tok}(\mathcal{A})$, and $C = \mathrm{Typ}(\mathcal{A})$. The relation $\Vdash_{\mathcal{A}}$ will have to be defined so that $\leq$, $\perp$, and $|$, are respected. This requirement leads to:

Definition 7.4. A *populated ontology* is a tuple $\widetilde{\mathcal{O}} = (\mathcal{A}, \leq, \perp, |)$ such that $\mathcal{A} = \langle X, C, \Vdash_{\mathcal{A}} \rangle$ is an information flow classification, $\mathcal{O} = (\mathcal{A}, \leq, \perp, |)$ is an ontology, and for all $x \in X$, and $c, d \in C$, we have:

(1) if $x \Vdash_{\mathcal{A}} c$, and $c \leq d$, then $x \Vdash_{\mathcal{A}} d$;

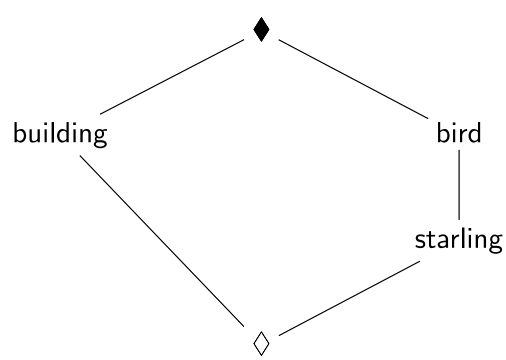
(2) if $x \Vdash_{\mathcal{A}} c$, and $c \perp d$, then $x \nVdash_{\mathcal{A}} d$;

(3) if $c|d$, then $x \Vdash_{\mathcal{A}} c$, or $x \Vdash_{\mathcal{A}} d$.

A populated ontology $\widetilde{\mathcal{O}} = (\mathcal{A}, \leq, \perp, |)$ having $\mathcal{A} = \langle X, C, \Vdash_{\mathcal{A}} \rangle$, determines a local logic $\mathfrak{L} = (\mathcal{A}, \vdash)$, whose theory $(C, \vdash)$, is given by the smallest regular theory (i.e. the smallest theory closed under Identity, Weakening, and Global Cut), such that for all $c, d \in C$, we have:

$$\begin{aligned} c \vdash d &\Leftrightarrow c \leq d \\ c, d \vdash &\Leftrightarrow c \perp d \\ \vdash c, d &\Leftrightarrow c|d \end{aligned} \tag{7.7}$$

Example 7.1. To get an idea of what these last relations mean, take the case of a reference ontology $\mathcal{O} = (C, R, \leq, \perp, |, \sigma)$ as in Kalfoglou and Schorlemmer (2003, §4), with a set of concepts $C = \{\mathsf{building}, \mathsf{bird}, \mathsf{starling}\}$, the relation $R = \{\mathsf{isHavenFor}\}$, arities $\sigma(\mathsf{isHavenFor}) = \langle \mathsf{building}, \mathsf{bird} \rangle$, where the partial order $\leq$, disjointness $\perp$, and coverage $|$, are defined by the following lattice:



(7.8)

where ♦ is the top and ♢ is the bottom of the lattice, i.e. $\mathsf{building} \perp \mathsf{bird}$ and $\mathsf{building}|\mathsf{bird}$. In this set up, we then have

$$\begin{aligned} \mathsf{building}, \mathsf{bird} &\vdash \\ \mathsf{starling} &\vdash \mathsf{bird} \\ &\vdash \mathsf{building}, \mathsf{bird} \end{aligned} \tag{7.9}$$

where the comma on the left-hand side has conjunctive force, whereas on the right-hand side it has disjunctive force. Thus, with respect to set of concepts C, the above constraints declare, respectively: nothing is both a building and a bird, all starlings are birds, and everything (here) is either a building or a bird.

With respect to the theory in question, the corresponding sequents are:

$$\langle\{\mathsf{bird},\mathsf{starling}\},\{\mathsf{building}\}\rangle$$
$$\langle\{\mathsf{building}\},\{\mathsf{bird},\mathsf{starling}\}\rangle$$
$$\langle\{\mathsf{bird}\},\{\mathsf{building},\mathsf{starling}\}\rangle$$

Then we have a classification in terms of the above sequents, as given by:

$\Vdash_{\mathcal{A}}$	building	bird	starling
$\langle\{\mathsf{bird},\mathsf{starling}\},\{\mathsf{building}\}\rangle$	0	1	1
$\langle\{\mathsf{building}\},\{\mathsf{bird},\mathsf{starling}\}\rangle$	1	0	0
$\langle\{\mathsf{bird}\},\{\mathsf{building},\mathsf{starling}\}\rangle$	0	1	0

Let the above set of sequents be denoted by X. Note how the sequents code the classification $\mathcal{A} = \langle X, C, \Vdash_{\mathcal{A}}\rangle$ whereby the left-hand sides of these indicate which columns contain '1' entries, and the right-hand sides indicate which columns contain '0' entries. Assuming that X consists of normal tokens, as in Kalfoglou and Schorlemmer (2003), we obtain a local logic $\mathcal{L} = (\mathcal{A}, \vdash)$ of the ontology $\mathcal{O}$. Given that $\mathcal{L}$ is a local logic on $\mathcal{A}$, we have by Barwise and Seligman (1997, Prop. 12.7), that $\mathcal{L} = \mathsf{Lg}(\mathcal{A})$; that is, with regards to Definition 6.6, $\mathcal{L}$ is the local logic generated by the classification $\mathcal{A}$. Associated to $\mathcal{O}$ is a local, populated ontology, as shown in Kalfoglou and Schorlemmer (2003, §4), to which we refer for details.

Example 7.2. In order to formalize semantic integration of a collection of agents in Channel Theory, Kalfoglou and Schorlemmer (2004) propose: i) modeling populated ontologies of agents by classifications; ii) defining the channel, its core, and infomorphisms between classifications; iii) defining a logic on the core of the channel; and, iv) distributing the logic to the sum of agent classifications to obtain the required theory for semantic interoperability within the channel. These steps give rise to a *global ontology* for two candidate agents, A_1 and A_2, requiring semantic integration. This commences with a distributed logic of a channel $\mathcal{C}$ connecting the classifications $\mathbf{A_1}$ and $\mathbf{A_2}$ that model the agents' populated ontologies $\widetilde{\mathcal{O}}_1$ and $\widetilde{\mathcal{O}}_2$, respectively:

$$\begin{tikzcd} & \mathbf{C} & \\ \mathbf{A}_1 \arrow[ur, "f_1"] \arrow[rr, "f"'] & & \mathbf{A}_2 \arrow[ul, "f_2"'] \end{tikzcd} \tag{7.10}$$

At the core of the channel $\mathcal{C}$, $\mathrm{Typ}(\mathbf{C})$ covers $\mathrm{Typ}(\mathbf{A}_1)$ and $\mathrm{Typ}(\mathbf{A}_2)$, while elements of $\mathrm{Tok}(\mathbf{C})$ connect tokens from $\mathrm{Tok}(\mathbf{A}_1)$ with those from $\mathrm{Tok}(\mathbf{A}_2)$. Effectively, the global ontology comes about when the logic on the core of the channel is distributed to the sum of classifications $\mathbf{A}_1 + \mathbf{A}_2$, for the total semantic integration of the combined events.

The structure of a typical ontology mapping may thus be seen as follows Kalfoglou and Schorlemmer (2003):

$$
\begin{tikzcd}
 & \mathcal{O}_{\mathrm{r}} \arrow[dl] \arrow[dr] & \\
\mathcal{O}_{\mathrm{loc}_1} \arrow[dr, dotted] & & \mathcal{O}_{\mathrm{loc}_2} \arrow[dl, dotted] \\
 & \mathcal{O}_{\mathrm{glob}} &
\end{tikzcd}
\tag{7.11}
$$

where $\mathcal{O}_{\mathrm{r}}$ is a reference ontology, $\mathcal{O}_{\mathrm{loc}_1}, \mathcal{O}_{\mathrm{loc}_2}$ are local ontologies, and $\mathcal{O}_{\mathrm{glob}}$ is a global ontology.

7.4 Quotient channel

Given an invariant $I = \langle \Sigma, R \rangle$ on a classification $\mathcal{A}$, the *quotient channel of* $\mathcal{A}$ *by* I is the limit of the distributed system depicted by

$$\mathcal{A} \xleftarrow{\tau_I} \mathcal{A}/I \xrightarrow{\tau_I} \mathcal{A}. \tag{7.12}$$

As a refinement of any other such channel, the quotient channel makes the following diagram commute

$$
\begin{tikzcd}
 & \mathbf{C} & \\
\mathcal{A} \arrow[ur, "g_1"] & \mathcal{A}/I \arrow[l, "\tau_I"'] \arrow[r, "\tau_I"] \arrow[u, "h"] & \mathcal{A} \arrow[ul, "g_2"']
\end{tikzcd}
\tag{7.13}
$$

Remark 7.3. An example is a hierarchially modular chain, where at each level of abstraction, the tokens can be inherited, and the resulting infomorphisms are created by systematically composing those from the levels below. As seen in Franklin and Patterson (2006) or Friedlander and Franklin (2008), the perception-to-memory relationships and actions of a typical LIDA semantic network architecture appear to fit into this pattern.

Remark 7.4. Ideas such as information flow, formal concepts, conceptual spaces, and local logics can be categorically unified when they are embraced within the abstract axiomatization of an *Institution* (Goguen and Burstall, 1992; Goguen, 2005a,b). This consists of a functor from an abstract category of 'Signatures' to a category of classifications that involves 'contexts' linked via the 'satisfaction relation' ($\Vdash$). As pointed out in Kent (2016), information flow is a particular case of FOL (which is thus one Institution), but the classification relation between Tokens and Types abstracts the Institution satisfaction relation between structures and sentences. For a further application of these concepts to ontologies, see Spivak and Kent (2012).

7.5 State spaces and projections

Recall that a state space is a classification S for which each token is of exactly one type, and where the types of the space are simply the states themselves. Here a is said to be *in state* σ if $a \Vdash_{\mathsf{S}} \sigma$. The space S is *complete* if every state is the state of some token.

Definition 7.5. A *projection* $f : \mathsf{S}_1 \rightrightarrows \mathsf{S}_2$ from a state space S_1, to a state space S_2 is given by a covariant pair of functions, such that for each token $a \in \mathrm{Tok}(\mathsf{S}_1)$, we have $f(\mathrm{state}_{\mathsf{S}_1}(a)) =$

$\mathrm{state}_{\mathsf{S}_2}(f(a))$. This amounts to the commutativity of the following diagram:

$$\begin{tikzcd} \mathrm{Typ}(\mathsf{S}_1) \arrow[r, "f"] & \mathrm{Typ}(\mathsf{S}_2) \\ \mathrm{Tok}(\mathsf{S}_1) \arrow[u, "\mathrm{state}_{\mathsf{S}_1}"] \arrow[r, "f"'] & \mathrm{Tok}(\mathsf{S}_2) \arrow[u, "\mathrm{state}_{\mathsf{S}_2}"'] \end{tikzcd} \tag{7.14}$$

The composition of projections is well-defined, and so too is the Cartesian product $\Pi_{i\in\mathcal{I}}\mathsf{S}_i$ of indexed state spaces with natural projections

$$\pi_{\mathsf{S}_i} : \Pi_{i\in\mathcal{I}}\mathsf{S}_i \rightrightarrows \mathsf{S}_i, \tag{7.15}$$

(see Barwise and Seligman (1997, §8.2)).

7.6 The Event Classification

Following the development of ideas in Barwise and Seligman (1997, Ch. 9), studies such as Kakuda and Kikuchi (2001) define the *Event Classification* $\mathsf{Evt}(\mathsf{S})$ associated with a state space S as follows:

(1) $\mathrm{Tok}(\mathsf{Evt}(\mathsf{S})) = \mathrm{Tok}(\mathsf{S})$;

(2) $\mathrm{Typ}(\mathsf{Evt}(\mathsf{S})) = \mathcal{P}(\mathrm{Typ}(\mathsf{S}))$;

(3) $s \Vdash_{\mathsf{Evt}(\mathsf{S})} \alpha$ is defined by $\mathsf{state}_s(s) \in \alpha$, for $s \in \mathrm{Tok}(\mathsf{Evt}(\mathsf{S}))$ and $\alpha \in \mathrm{Typ}(\mathsf{Evt}(\mathsf{S}))$;

where as before, $\mathcal{P}(\cdot)$ indicates the power set.

Briefly recapping, this says that the space of events $\mathrm{Evt}(\mathsf{S})$ associated to S has as its tokens the tokens of S, and its types are arbitrary sets of sets of states of S. The classification relation $s \Vdash_{\mathrm{Evt}(\mathsf{S})} \alpha$ above is equivalent to $\mathrm{state}_{\mathsf{S}}(s) \in \alpha$. Following Barwise and Seligman (1997, Prop. 8.17), given state spaces S_1 and S_2, the following are equivalent:

(1) $f : \mathsf{S}_1 \rightrightarrows \mathsf{S}_2$ is a projection;

(2) $\mathrm{Evt}(f) : \mathrm{Evt}(\mathsf{S}_2) \rightleftarrows \mathrm{Evt}(\mathsf{S}_1)$ is an infomorphism.

In fact, for any state space S, the classification $\mathrm{Evt}(\mathsf{S})$ is a Boolean classification in which the operations of taking intersection, union, and complement are here conjunction, disjunction, and negation, respectively (Barwise and Seligman, 1997, Prop. 8.18).

Definition 7.6. Let S be a state space. The *local logic generated by* S, denoted $\mathsf{Lg}(\mathsf{S})$, has classification $\mathrm{Evt}(\mathsf{S})$, regular theory $\mathsf{Th}(\mathsf{S})$, and all of its tokens are normal.

For further relationships see Barwise and Seligman (1997, 12.1 - 12.2).

Remark 7.5. Taking $\mathsf{K} = [0, 1]$, the evaluation relation

$$\Vdash_{\mathsf{Evt}(\mathsf{S})} : \mathrm{Tok}(\mathsf{Evt}(\mathsf{S})) \times \mathrm{Typ}(\mathsf{Evt}(\mathsf{S})) \longrightarrow [0, 1] \tag{7.16}$$

together with (logic) infomorphisms between event classifications, may be compared with the concept of a *perceptual strategy* as described in Hoffman, Singh and Prakash (2015).

7.7 State space systems

Considering the above ingredients we now seek a unifying principle that characterizes the state space model and provides a suitable information channel. This motivates starting with:

Definition 7.7. A *state-space system* consists of an indexed family $\mathcal{S} = \{f_i : \mathsf{S} \rightrightarrows \mathsf{S}_i\}_{i \in \mathcal{I}}$ of state-space projections with a common domain S, called the *core* of $\mathcal{S}$, to state spaces S_i (for $i \in \mathcal{I}$); S_i is called the *ith component space of* $\mathcal{S}$.

We now consider an "event" Evt as a functor that transforms a state-space system into an information channel. Taking a pair of state spaces as an example, we first take projections:

$$\begin{tikzcd} & \mathsf{S} \arrow[dl, "f_i"'] \arrow[dr, "f_j"] & \\ \mathsf{S}_i & & \mathsf{S}_j \end{tikzcd} \tag{7.17}$$

Next, applying the functor Evt to this diagram yields a family of infomorphisms with a commom domain $\mathrm{Evt}(\mathsf{S})$, yielding an information channel:

$$\begin{tikzcd} & \mathrm{Evt}(\mathsf{S}) & \\ \mathrm{Evt}(\mathsf{S}_i) \arrow[ur, "\mathrm{Evt}(f_i)"] & & \mathrm{Evt}(\mathsf{S}_j) \arrow[ul, "\mathsf{Evt}(f_j)"'] \end{tikzcd} \tag{7.18}$$

State space addition produces a further commuting diagram, where for ease of notation, we write σ_i for $\sigma_{\mathrm{Evt}(\mathsf{S}_i)}$, and simply f for $\sum_{k \in \mathcal{I}} \mathrm{Evt}(f_k)$:

$$\begin{tikzcd} & \mathrm{Evt}(\mathsf{S}) & \\ \mathrm{Evt}(\mathsf{S}_\mathrm{i}) \arrow[ur, "\mathrm{Evt}(f_i)"] \arrow[r, "\sigma_i"'] & \sum_{k \in \mathcal{I}} \mathrm{Evt}(\mathsf{S}_\mathrm{k}) \arrow[u, "f"] & \mathrm{Evt}(\mathsf{S}_\mathrm{j}) \arrow[ul, "\mathrm{Evt}(f_j)"'] \arrow[l, "\sigma_j"] \end{tikzcd} \tag{7.19}$$

Example 7.3. Following Kakuda and Kikuchi (2001, §4), let T be a regular theory, and let S be a state space. A *medium system* denoted $D := \langle \mathsf{D}, \mathsf{N}, f, p \rangle$ between T and S, consists of the following:

(1) a state space D;

(2) a subset N of $\mathrm{Tok}(D)$;

(3) an infomorphism $f : \mathsf{Cl}(T_{\mathrm{Typ}(T)}) \rightleftarrows \mathsf{Evt}(\mathsf{D})$;

(4) a projection $p : \mathsf{D} \rightrightarrows \mathsf{S}$;

where $\langle \mathrm{Tok}(\mathsf{D}), \mathrm{Typ}(T), \vdash_{\mathrm{T}}, \mathsf{N}, \mathrm{Typ}(D), \mathsf{state}_{\mathrm{D}}, \overrightarrow{f}, \overleftarrow{f} \rangle$ forms a functional scheme. Here D is called the *medium space of* D. Kakuda and Kikuchi (2001, §4) define an information channel:

$$\begin{tikzcd} & \mathsf{Evt}(\mathsf{D}) & \\ \mathsf{Cl}(T_{\mathrm{Typ(T)}}) \arrow[ur, "f"] & & \mathsf{Evt}(\mathsf{S}) \arrow[ul, "\mathsf{Evt}(p)"'] \end{tikzcd} \tag{7.20}$$

through $\mathsf{Evt}(\mathsf{D})$ to $\mathsf{Evt}(\mathsf{S})$.

Example 7.4. Sakahara and Sato (2008, 2011) employ the concept of the core of a binary channel, when realized as a classification, in order to describe (modal) logical relationships holding for an agent whose cognition is determined by some regular theory as defined in §6.7 above (cf. Barwise, 1997). Consider separate source ($\mathcal{A}$) and target ($\mathcal{B}$) classifications, and represent the agent's knowledge by a regular theory $T = \langle \Sigma, \vdash \rangle$. The idea is to construct a set of *possible and realizable states* in several steps:

(1) For a source classification $\mathcal{A}$, and a target classification $\mathcal{B}$, let $\Omega_{\langle \mathcal{A},\mathcal{B} \rangle}$ denote the set of all partitions of $\mathrm{Typ}(\mathcal{A}) \cup \mathrm{Typ}(\mathcal{B})$, called *the set of states generated by $\mathcal{A}$ and $\mathcal{B}$.*

(2) The set of *realizable states generated by $\mathcal{A}$ and $\mathcal{B}$* is given by

$$\Omega^{R}_{\langle \mathcal{A},\mathcal{B} \rangle} := \{\langle \Theta, \Lambda \rangle \in \Omega_{\langle \mathcal{A},\mathcal{B} \rangle} : \exists a \in \mathcal{A},\ \mathrm{Typ(a)} \subseteq \Theta \text{ and, } \mathrm{Typ^c(a)} \subseteq \Lambda\}, \tag{7.21}$$

where the notation $\mathrm{Typ^c(a)}$ indicates the complement, i.e. everything not in $\mathrm{Typ(a)}$.

(3) The set of *impossible states under the theory T* is given by

$$\Omega^{IP}_{\langle \mathcal{A},\mathcal{B}|T \rangle} := \{\langle \Theta, \Lambda \rangle \in \Omega_{\langle \mathcal{A},\mathcal{B} \rangle} : \Theta \vdash_T \Lambda\}. \tag{7.22}$$

The "impossibility" here is that Ω constrains a Λ with which it is disjoint.

(4) *The possible states under the theory T* is then

$$\Omega^{P}_{\langle \mathcal{A},\mathcal{B}|T \rangle} = \Omega_{\langle \mathcal{A},\mathcal{B} \rangle} \backslash \Omega^{IP}_{\langle \mathcal{A},\mathcal{B}|T \rangle}. \tag{7.23}$$

(5) *The possible and realizable states under the theory T* is thus

$$\Omega^{PR}_{\langle \mathcal{A},\mathcal{B}|T \rangle} = \Omega^{P}_{\langle \mathcal{A},\mathcal{B}|T \rangle} \cap \Omega^{R}_{\langle \mathcal{A},\mathcal{B} \rangle}. \tag{7.24}$$

The *cognizance classification* $\mathfrak{C}_{\langle \mathcal{A},\mathcal{B},T \rangle}$ generated by $\mathcal{A}, \mathcal{B}$ and T is then given by

$$\mathfrak{C}_{\langle \mathcal{A},\mathcal{B},T \rangle} := \left\langle \Omega^{PR}_{\langle \mathcal{A},\mathcal{B}|T \rangle}, \mathrm{Typ}(\mathcal{A}) \cup \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathfrak{C}_{\langle \mathcal{A},\mathcal{B},\mathrm{T} \rangle}} \right\rangle \tag{7.25}$$

where the relation $\Vdash_{\mathfrak{C}_{\langle \mathcal{A},\mathcal{B},T \rangle}}$ is defined as $\langle \Theta, \Lambda \rangle \Vdash_{\mathfrak{C}_{\langle \mathcal{A},\mathcal{B},T \rangle}} \alpha$, if and only if $\alpha \in \Theta$, i.e. the choice of Λ can be arbitrary provided it is disjoint from Θ.

7.8 On comparing and combining the Shannon Theory of Information with Channel Theory

Barwise (1997) recalls *Shannon's Inverse Relation* between possibilities and information, basically saying that eliminating possibilities from consideration amounts to increasing one's information and vice-versa. That relationship is fundamental to Dretske's original goal of developing a semantic theory of information based on possibilities (Dretske, 1981, 2000). Though very general as a quantitative theory of communication flow, the original Shannon theory had largely overlooked the question of semantic content. In any Dretske-type theory, the basis of semantic content is in the world, i.e. in the events or situations that signals or states carry information about. By showing how local logics are connected by information networks, Channel Theory provides a general qualitative theory of information flow in this context. 'Channels' in the theory are more general than the traditional idea of the Shannon communication channels (Cover and Thomas, 2006). In Shannon's theory, information flow in a channel is defined in terms of reduction of uncertainty about the *type* of event that will occur; it says nothing about the semantics of any specific bit, i.e. about any specific token. Channel Theory specifically concerns particular tokens x and statements of the form "x is an A." Allwein (2004); Allwein, Moskowitz and Chang (2004) create a synthesis of Shannon's quantitative theory with the Barwise-Seligman qualitative theory to address the question of how specific objects, situations and events carry information about each other.

A classification of possible outcomes of events starts with a probability space $\mathcal{P} = \langle \Omega, \Sigma, \mu \rangle$, where Ω denotes the set of possible outcomes, Σ is a σ-algebra on Ω whose members represent events, and μ is a probability measure on σ representing the probability of an event having or being associated with a particular outcome.[¶] We define a classification $\mathrm{Tok}(\mathcal{P}) = \Omega$, $\mathrm{Typ}(\mathcal{P}) = \Sigma$, and let $\omega \Vdash_{\mathcal{P}} e$, if and only if $\omega \in e$. In this context, an infomorphism between probability spaces is topologically a continuous map (Seligman, 2009).

To see how the basic ontology of the Shannon theory can be conveniently embedded in that of Channel theory, note that the former's basic unit of information is some tuple of a binary relation. The relation is restricted to be of the form $x \Vdash V$, where $\Vdash$ is regarded as a function, and V is the value of the Token x. But closer inspection reveals this characterizes a state space in which V is a state and tokens are ignored. In Channel Theory there are states, but tokens are not ignored and types are values, as in §6.1. States may be amalgamated to form Events as in §7.6 and §7.7. Channel theory permits this by firstly preserving the tokens, and then replacing states with types, whose events are also types: for some event E, $x \Vdash E$, just when $x \Vdash s$, for some state $s \in E$ (Allwein, Moskowitz and Chang, 2004). This is basically the embedding of the one theory into the other, and the two theories together admit a certain generalization as follows.

The presence of a sequent in an information channel (as outlined in §6.8) effectively represents a logical gate, and this kind of structure can be seen as more general than a Markov structure (cf. the Kolmogorov axioms in Allwein, 2004; Allwein, Moskowitz and Chang, 2004), since sequents enable the information flow to simultaneously support a flow of reasoning. Specifically, probabilities can be assigned to sequents in $\mathcal{A}$ as follows. Suppose we have:

$$M \Vdash_{\mathcal{A}} N \qquad \forall x (x \Vdash_{\mathcal{A}} M \Rightarrow x \Vdash_{\mathcal{A}} N), \tag{7.26}$$

then the sequent relation $\vdash$ can be weakened by removing $\forall$, and instead stating that for any x, we have a probability $x \Vdash N$, given $x \Vdash M$; that is, the probability that x satisfies N given it satisfies M. This is clearly a conditional probability, so one defines:

$$M \Vdash_{\mathcal{A}}^{\mathbf{P}} N := \mathbf{P}(M|N). \tag{7.27}$$

Thus when the sequent's conditional probability is p, say, we have $M \Vdash_{\mathcal{A}}^{p} N$. *A priori*, one must have $x \Vdash_{\mathcal{A}} M$ in order to apply $M \Vdash_{\mathcal{A}} N$ in a argument. The probability of the former holding in $\mathcal{A}$, is $\mathbf{P}(M)$. Then $x \Vdash_{\mathcal{A}} N$ follows from the rule $\mathbf{P}(M) \cdot M \Vdash_{\mathcal{A}}^{\mathbf{P}} N$. Probability axioms for a Countable Classical Propositional Logic are developed in Allwein (2004) (cf. Allwein, Moskowitz and Chang, 2004) to which we refer for details. Note that information flow in distributed systems can be interpreted dynamically; this amounts to causation in an informational context, consistent

[¶]Recall that a *σ-algebra* over Ω is a set Σ of subsets of Ω, such that $\emptyset \in \Sigma$, $\Omega - e \in \Sigma$, for each $e \in \Sigma$, and $\bigcup E \in \Sigma$, for each countable set $E \subseteq \Sigma$. μ is a *probability measure* on Σ, if and only if it satisfies the Kolmogorov axioms: $\mu(\emptyset) = 0$, $\mu(\Omega - e) = 1 - \mu(e)$, and $\mu(\bigcup E) = \sum_{e \in E} \mu(e)$ if E is countable, and $\mu(e_1 \cap e_2) = 0$, for all $e_1 \neq e_2 \in E$.

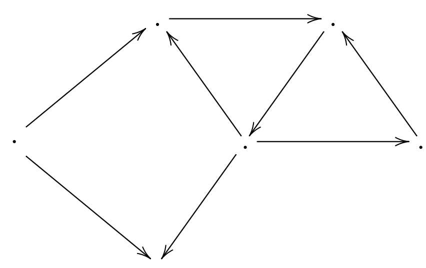
with the Dretskean nature of the theory. In this respect, the relations between information theory and logic are also conducive to understanding certain relations between causation and computation (Collier, 2011; Seligman, 2009).

8 Colimits for piecing together local-to-distributed information: Application to information flow

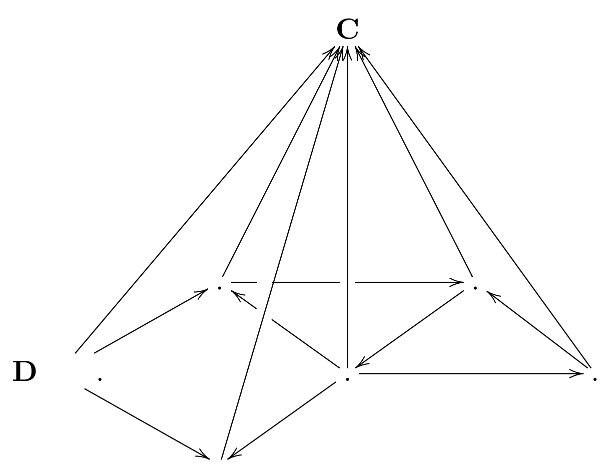
8.1 Cocones and Colimits

To a category theorist, minimal covers as described for Distributed Systems in §7.2 are familiar as *colimits*, where the channel core **C** represents the vertex of a cocone (as was introduced in §6.4; more formally, see e.g. Awodey (2010, Ch. 5)), and, as pointed out in Barwise and Seligman (1997, §2.1), the existence of colimits of Chu spaces (classifications) was established in Barr (1979, 1991). Let us briefly recall this concept with some graphic intuition behind the idea following Baianu et al. (2006); Brown and Porter (2003).

The "input data" for a colimit is a *diagram* **D**, i.e. a collection of some objects in a category $\mathfrak{C}$, together with some arrows between them, as depicted by:

$\mathbf{D} =$  (8.1)

This generalizes our use of a directed graph in §6.4 by allowing the "vertices" and "edges" of **D** to be objects and morphisms of an arbitrary category. Next we need 'functional controls' comprising a *cocone with base* **D** *and vertex an object* **C** *in* $\mathfrak{C}$,



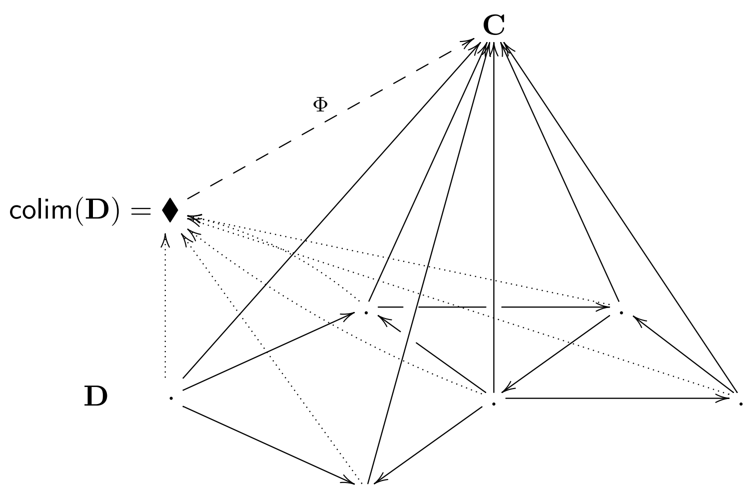
(8.2)

such that each of the triangular faces of the cocone is commutative. The "output" from **D** will be an object $\mathsf{colim}(\mathbf{D})$ in our category $\mathfrak{C}$ defined by a special *colimit cocone*, such that any cocone on $\mathbf{D}$ factors *uniquely* through the colimit cocone. Effectively, the commutativity condition on the cocone induces, in the colimit, an interaction of images from different parts of the diagram $\mathbf{D}$. The uniqueness condition makes the colimit the optimal solution to this factorisation problem.

Let us set

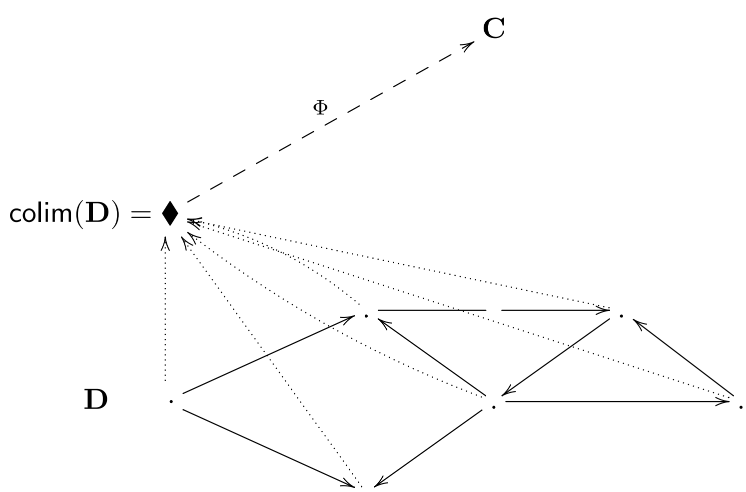
$$\blacklozenge = \mathsf{colim}(\mathbf{D}) \tag{8.3}$$

where the dotted arrows in the diagram below represent new morphisms which combine to make the colimit cocone:



(8.4)

and for which the broken arrow Φ is constructed by requiring commutativity for all of the triangular faces of the combined diagram. Next, stripping away the 'old' cocone results in a factorisation of the cocone via the colimit:



(8.5)

Intuitively, the process can be seen as follows. The object $\mathsf{colim}(\mathbf{D})$ is pieced together from the diagram $\mathbf{D}$ by means of the colimit cocone. From beyond $\mathbf{D}$, an arbitrary object $\mathbf{C}$ in $\mathfrak{C}$ 'sees' $\mathbf{D}$ as mediated through its colimit. This means that if $\mathbf{C}$ is going to interact with all of $\mathbf{D}$, then it does so via $\mathsf{colim}(\mathbf{D})$. The colimit cocone can be thought of as a kind of program: given any cocone on

$\mathbf{D}$ with vertex $\mathbf{C}$, the output will be a morphism

$$\Phi : \mathsf{colim}(\mathbf{D}) \to \mathbf{C} \tag{8.6}$$

as constructed from the other data. [‖]

Example 8.1. Brown and Porter (2003) provide an analogy comparing colimits with how an email message can be relayed. Suppose E denotes some email document. This is to be sent via a server S, which decomposes E into numerous parts E_i ($i \in \mathcal{I}$, an indexing set), and labels each part E_i, so it becomes E_i'. These labelled parts E_i' are then sent to a number of servers S_i, which then relay these messages as newly labelled messages E_i'' to a server S_C, for the receiver C. The server S_C then combines the E_i'' to produce the recovered message M_E at C. Breaking the message down and routing it through the S_i appears arbitrary, but the system is designed such that M_E is independent of all choices made at each step of the process.

Many other illustrative examples applying colimits to computer science, social systems, and neuroscience, can be seen in Baianu et al. (2006); Ehresmann and Vanbremeersch (2007); Healy and Caudell (2006); Healy (2010); Porter (1994).

8.2 Coordinated channels in ontologies

Here we describe how the colimit concept features in semantic integration within an information channel. Suppose we have two prospectively interoperating agents A_1, A_2, with each agent A_i ($i = 1, 2$) having its knowledge represented according to its own conceptualization, as specified in relationship to its ontology $\mathcal{O}_i$, respectively. This means that a concept of $\mathcal{O}_1$ will, *a priori*, be considered semantically distinct from $\mathcal{O}_2$, even if they are equivalent syntactically. However, the behavior of the agents can provide evidence for a meaning common to A_1 and A_2. Let us us assume that the agents' ontologies are not open to a third-party inspection. Kalfoglou and Schorlemmer (2004) use a channel to coordinate the populated ontologies (cf. §7.3) $\widetilde{\mathcal{O}}_1, \widetilde{\mathcal{O}}_2$ by capturing the degree of participation of each agent in communicative behaviors. Specifically, i) agent A_i attempts to "explain" a subset of its concepts to other agents, and ii) other agents exchange with A_i some of their own tokens, thus increasing the set of tokens originally available to A_i.

To see how this degree of participation can be captured by Channel Theory, Kalfoglou and Schorlemmer (2004) introduce classifications $\mathbf{A}_i = \langle \mathrm{Tok}(\mathbf{A}_\mathrm{i}), \mathrm{Typ}(\mathbf{A}_\mathrm{i}), \Vdash_{\mathbf{A}_\mathrm{i}} \rangle$, corresponding to the agents A_i, respectively, along with subclassifications $\mathbf{A}_i' = \langle \mathrm{Tok}(\mathbf{A}_\mathrm{i}'), \mathrm{Typ}(\mathbf{A}_\mathrm{i}'), \Vdash_{\mathbf{A}_\mathrm{i}'} \rangle$, and infomorphisms $g_i : \mathbf{A}_i' \longrightarrow \mathbf{A}_i$, for which functions $\hat{g}_i$ and $\check{g}_i$ are the inclusions $\mathrm{Typ}(\mathbf{A}_\mathrm{i}') \subseteq \mathrm{Typ}(\mathbf{A}_\mathrm{i})$ and $\mathrm{Tok}(\mathbf{A}_\mathrm{i}') \subseteq \mathrm{Tok}(\mathbf{A}_\mathrm{i})$, respectively. It is from the subclassifications $\mathbf{A}_i'$ arising from the interactions that coordination is established. Thus we have the following information channel with core (i.e. cocone) $\mathbf{C}'$:

$$\begin{tikzcd}
 & & \mathbf{C}' & & \\
\mathbf{A}_1 & \mathbf{A}_1' \arrow[l, "g_1"'] \arrow[ur, "f_1"] & & \mathbf{A}_2' \arrow[ul, "f_2"'] \arrow[r, "g_2"'] & \mathbf{A}_2
\end{tikzcd} \tag{8.7}$$

[‖]The diagrams included in (8.1)–(8.5) are reproduced from Baianu et al. (2006); Brown and Porter (2003), with permission from R. Brown

The optimal coordinated channel that captures semantic integration achieved by the agents is then represented by the colimit $\mathcal{C}' = \mathsf{colim}\{\mathbf{A}'_1 \leftarrow \mathbf{S} \rightarrow \mathbf{A}'_2\}$ of the diagram linking the subclassifications that model the agents' participation in the interoperation:

$$\begin{tikzcd}
 & & \mathbf{C}' & & \\
\mathbf{A}_1 & \mathbf{A}'_1 \arrow[l, "g_1"'] \arrow[ur, "f_1"] & \mathbf{S} \arrow[l, "h_1"] \arrow[r, "h_2"'] & \mathbf{A}'_2 \arrow[r, "g_2"'] \arrow[ul, "f_2"'] & \mathbf{A}_2
\end{tikzcd} \tag{8.8}$$

"Optimality" here means that every other channel induces a map to $\mathcal{C}'$ when commutativity is required. Similar techniques using information channels, along with colimits, are developed in the framework of structure and language for logical environments in Kent (2016).

9 Conclusion

Category theory is in essence a theory of dualities. In this Part I, we have surveyed the techniques of Chu spaces and Channel Theory in a semantically based, ontological framework, which will be applied in Part II (Fields and Glazebrook, 2018) to fundamental aspects of cognition, and in particular to the study of visual object identification, by emphasizing the role of concepts and processes representable as category-theoretic duals in cognitive processing. The roles of complementary information flows at all scales, from on-center/off-surround networks to the dorsal and ventral attention systems to the interplay of memory and prediction that constructs object histories, exemplify such duality. Our motivation has been driven by the fact that, by taking object identities and object persistence for granted, AI systems have largely neglected the problem of object re-identification that lies at the heart of the frame problem (Fields, 2013, 2016). Taking this problem and the dual organization required to solve it into account suggests reconceptualizations of learning and memory as overarching dual processes, towards which Part II represents an initial foray. It is, finally, possible that the conceptual ground which we have covered in this Part I could be further supplemented by some related techniques of higher dimensional algebra (Brown, Higgins and Sivera, 2011) and those of n-categories (Leinster, 2004), topics which remain for further investigation.

Acknowledgment

The authors wish to thank several anonymous referees for their constructive criticism and helpful comments which assisted improving the overall presentation. JFG also wishes to thank Professors Charles Delman and Timothy Porter for interest in this work, and for various discussions.

• The authors report no conflict of interest involved in this work.

References

Abramsky, S. (2012). Big toy models: Representing physical systems as Chu spaces. *Synthese, 186*, 697–718.

Allwein, G. (2004). A qualititative framework for Shannon Information theories. In: NSPW '04 Proceedings of 2004 Workshop on New Security Paradigms. New York: ACM (pp. 23-31).

Allwein, G., Yang, Y. & Harrison, W. L. (2011). Qualitative decision theory via Channel Theory. *Logic Logical Phil, 20*, 1–30.

Allwein, G., Moskowitz, I.S. & Chang, L.-W. (2004). A new framework for Shannon information theory. Technical Report A801024. Washington, DC: Naval Research Laboratory.

Awodey, S. (2010). Category Theory. Oxford Logic Guides 62. Oxford, UK: Oxford University Press.

Baars, B. J. & Franklin, S. (2003). How conscious experience and working memory interact. *Trends Cogn. Sci., 7*, 166–172.

Baars, B. J., Franklin, S. & Ramsoy, T. Z. (2013). Global workspace dynamics: Cortical "binding and propagation" enables conscious contents. *Front. Psychol., 4*, 200.

Baianu, I. C., Brown, R., Georgescu, G. & Glazebrook, J. F. (2006). Complex nonlinear biodynamics in categories, higher dimensional algebra and Łukasiewicz-Mosil Topos: Transformations of neuronal, genetic and neoplastic networks. *Axiomathes,16*, 65–122.

Barr, M. (1979). *-Autonomous categories (with Appendix by Po-Hsiang Chu). *Lecture Notes in Mathematics, 752*. Heidelberg: Springer.

Barr, M. (1991). *-Autonomous categories and linear logic. *Math. Struct. Comp. Sci, 1*, 159–178.

Barwise, J. & Seligman, J. (1997). Information Flow: The Logic of Distributed Systems. Cambridge Tracts in Theoretical Computer Science 44. Cambridge, UK: Cambridge University Press.

Barwise, J. (1997). Information and Impossibilities. *Notre Dame J. Formal Logic, 38*, 488–515.

Barwise, J. (1999). State spaces, local logics, and non-monotonicity. In: Moss, L. et al. (Eds.) Logic, Language and Computation, Vol 2. (London, 1996), (CSLI Lecture Notes, 96) Stanford, CA: CSLI Publications (pp. 1–20).

Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Cambridge, MA: Bradford Books, MIT Press.

Berners-Lee, T., Hendler, J. & Lassila, O (2001) The Semantic Web. *Scientific American*, 284(5), 34–43.

Brentano, F. (1981). *The Theory of Categories* (Chisholm, R., and Guterman, N., translators). The Hague: Nijhoff.

Brown, R. & Porter, T. (2003). Category theory and higher dimensional algebra: Potential descriptive tools in neuroscience. In: Singh, N. (Ed.) Proceedings of the International Conference on Theoretical Neurobiology, Delhi, February 2003, National Brain Research Centre (pp. 80–92).

Brown, R., Higgins, P. J. & Sivera, R. (2011). *Nonabelian Algebraic Topology*. (Tracts in Mathematics, 15). Zürich: European Mathematical Society (EMS).

Caspard, N. & Monjardet, B. (2003). The lattices of closure systems, closure operators, and implicational systems on a finite set: A survey. *Disc. Appl. Math., 127*, 241–269.

Collier, J. D. (2011). Information, causation and computation. In: Dodig-Crnkovic, G. & Burgin, M.,(Eds.) Information and Computation: Essays on Scientific and Philosophical Foundations of Information and Computation. Singapore: World Scientific (pp. 89–105).

Cordier, J.-M. & Porter, T. (1989). *Shape theory: Categorical Methods of Approximation.* Chichester, UK: Ellis Horwood.

Cover, T. M. & Thomas, J. A. (2006). *Elements of Information Theory.* New York: Wiley.

Dawson, M. R. W. (2005). *Connectionism: A Hands-On Approach.* Malden, MA: Blackwell.

Dayan, P., Hinton, G. E., Neal, R. M. & Zemel, R. S. (1995). The Helmholtz machine, *Neural Comp., 7*, 1022–1037.

Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79*, 1–37.

Dehaene, S. & Changeux, J. P. (2004). Neural mechanisms for access to consciousness. In: Gazzaniga, M. S. (Ed.) The Cognitive Neurosciences, 3rd Edn. Cambridge, MA: MIT Press (pp. 1145–1157).

Dehaene, S., Charles, L., King, J.-R. & Marti, S. (2014). Toward a computational theory of conscious processing. *Curr. Opin. Neurobiol., 25*, 76–84.

Dowker, C. H. (1952). Homology groups of relations. *Ann. Math.,56*, 84–95.

Dretske, F. (1981). *Knowledge and the Flow of Information.* Cambridge MA: MIT Press.

Dretske, F. (2000). *Perception, Knowledge, and Belief: Selected Essays.* New York: Cambridge University Press.

Ehresmann, A. C. & Vanbremeersch, J.-P. (2007). *Memory Evolutive Systems; Hierarchy, Emergence, Cognition (Studies in Multidisciplinarity).* New York: Elsevier.

Eilenberg, S. & Mac Lane, S. (1945). Relations between homology and homotopy groups. *Ann. Math., 46*, 480–509.

Fields, C. & Glazebrook, J. F. (2018). A mosaic of Chu spaces and Channel Theory II: Applications to object identification and mereological complexity. *J. Exper. Theor. Artif. Intell.*, in press.

Fields, C. (1989). Consequences of nonclassical measurement for the algorithmic description of continuous dynamical systems. *J. Exper. Theor. Artif. Intell., 1*, 171–178.

Fields, C. (2012). The very same thing: extending the object token concept to incorporate causal constraints on individual identity. *Adv. Cogn. Psychol., 8*, 234–247.

Fields, C. (2013). How humans solve the frame problem. *J. Exper. Theor. Artif. Intell., 25*, 441–456.

Fields, C. (2016). Visual re-identification of individual objects: A core problem for organisms and AI. *Cogn. Proc., 17*, 1–13.

Franklin, S. & Patterson, F. G. J. Jr. (2006). The LIDA architecture: adding new modes of learning to an intelligent, autonomous, software agent. *IDPT-2006 Proceedings (Integrated Design and Process Technology).* Society for Design and Process Science (pp. 1–8).

Franklin, S., Madl, T., D′Mello, S. & Snaider, J. (2014). LIDA: a systems-level architecture for cognition, emotion and learning. *IEEE Trans. Auton. Mental Devel., 6*, 19–41.

Friedlander, D. & Franklin, S. (2008). LIDA and a theory of mind. *Front. Artif. Intell. Applic., 171*, 137–148.

Friedman, G. (2012). Survey article: An elementary illustrated introduction to simplicial sets. *Rocky Mountain J. Math., 42*, 353–423.

Friston, K. J. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci., 11*, 127–138.

Ganter, B., Wille, R. & Franzke, C. (1999.) *Formal Concept Analysis: Mathematical Foundations, Vol 7.* Berlin: Springer.

Goerss, P. G. & Jardine, J. F. (1999). *Simplicial Homotopy Theory* (Progress in Mathematics, Vol 174). Basel: Birkhäuser.

Goguen, J. (2005a). Information integration in institutions. Proposed for: Moss, L. (Ed.) *Thinking Logically: A Memorial Volume for Jon Barwise.* Bloomington IN: Indiana University Press (https://cseweb.ucsd.edu/ goguen/pps/ifi04.pdf).

Goguen, J. (2005b). What is a concept? In: Dau, F. & Mungier, M.-L. (Eds.) Proceedings, 13th Conference on Conceptual Structures (Lecture Notes in Artificial Intelligence, vol 3596). Kassel, Germany: Springer (pp. 52–77).

Goguen, J. & Burstall, R. (1992). Institutions: Abstract model theory for specification and programming. *J. Assoc. Comp. Mach., 39*, 95–146.

Gratus, J. & Porter, T. (2006). A spatial view of information. *Theor. Comp. Sci., 365*, 206–215.

Gratus, J. & Porter, T. (2005a). A geometry of information, I: Nerves, posets and differential forms. In: Kopperman, R. et al. (Eds.) Spatial Representation: Discrete vs. Continuous Computational Models, Dagstuhl Seminar Proceedings 04351, IBFI 2005, Schloss Dagstuhl, Germany.

Gratus, J. & Porter, T. (2005b). A geometry of information, II: Sorkin models and biextensional collapse.In: Kopperman, R. et al. (Eds.) Spatial Representation: Discrete vs. Continuous Computational Models, Dagstuhl Seminar Proceedings 04351, IBFI 2005, Schloss Dagstuhl, Germany.

Grossberg, S. (2013). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks, 37*, 1–47.

Healy, M. J. & Caudell, T. P. (2006). Ontologies and worlds in category theory: implications for neural systems. *Axiomathes,16*, 165–214.

Healy, M. J. (2010). Category theory as a mathematics for formalizing ontologies. In: Poli, R., Healey,M. & Kameas A. (Eds.) *Theory and Applications of Ontology: Computer Applications*. Berlin: Springer (pp. 487–510).

Hitzler, P. & Zhang, G.-Q. (2004). A Cartesian closed category of approximating concepts. Proc. 12th Internat. Conf. on Conceptual Structures, ICCS 2004, Huntsville, AL, July 2004. *(*Lecture Notes in Artificial Intelligence, 3127) Berlin: Springer (pp. 170–185).

Hitzler, P., Hölldobler, S. & Seda, A. K. (2004) Logic programs and connectionist networks. J. Appl. Logic, 2, 245–272.

Hoffman, D. D., Singh, M. & Prakash, C. (2015). The interface theory of perception. *Psychonom. Bull. Rev., 22*, 1480–1506.

Husserl, E. (1970). *Logical Investigations.* London: Routledge and Keagan Paul.

Kahneman, D., Triesman, A. & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cogn. Psychol., 24*, 175–219.

Kakuda, Y. & Kikuchi, M. (2001). Abstract design theory. *Ann. Japan Assoc. Phil. Sci., 10*(3), 19–35.

Kalfoglou, Y. & Schorlemmer, M. (2004). Formal support for representing and automating semantic interoperability. *The Semantic Web: Research and Applications. ESWS 2004, Heraklion, Crete* (Lecture Notes in Computer Science) Berlin: Springer (pp. 45–60).

Kalfoglou, Y. & Schorlemmer, M. (2003). IF-Map: An ontology-mapping method based on information-flow theory. *J. Data Semantics I* (Lecture Notes in Computer Science). Berlin: Springer (pp. 107–127).

Keifer, M. & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex, 7*, 805–825.

Kent, R. E. (2016). Information flow in logical environments. Preprint arXiv:1603.03475v1[cs.LO].

Kikuchi, M., Nagasaka, I., Toyoda, S. & Kitamura, S. (2003) A mathematical model of interactions in artifact environment. Proceedings of SICE Annual Conference 2003 (pp. 2085–2090).

Krötzsch, M., Hitzler, P. & Zhang, G.-Q. (2005) Morphisms in context. In: Dau, F., Mugnier, M.–L. & Stumme,G. (Eds.) ICCS 2005, LNAI 3596. Berlin: Springer (pp. 223–237).

Leinster, T. (2004). *Higher Operads, Higher Categories.* (London Mathematical Society Lecture Note Series Vol 298). Cambridge, UK: Cambridge University Press.

Leitgeb, H. (2005). Interpreted dynamical systems and qualitative laws: From neural networks to evolutionary systems. *Synthese, 146*, 189–202.

Mac Lane, S. (1971). *Categories for the Working Mathematician.* New York: Springer.

Maia, T. V. & Cleeremans, A. (2005) Consciousness: Converging insights from connectionist modeling and neuroscience. *Trends Cogn. Sci., 9*, 397–404.

Martin, A. (2007). The representation of object concepts in the brain. *Ann. Rev. Psychol., 58*, 25–45.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In: Oakford, M. & Chater, N. (Eds.) *Rational Models of Cognition.* Oxford, UK: Oxford University Press (pp. 21–53).

Nauck, D., Klawonn, F. & Kruse, R. (2003). *Neuronale Netze und Fuzzy-Systeme.* Braunschweig, Wiesbaden: Vieweg.

Nielsen M. A. & Chaung, I. L. (2000). *Quantum Computation and Quantum Information.* Cambridge UK: Cambridge University Press.

Poli, R. (2001). The basic problem of the theory of levels in reality. *Axiomathes, 12*, 261–283.

Poli, R. (2006). The theory of levels of reality and the difference between simple and tangled hierarchies. In: Minati, G., Pessa, E. & Abram, M. (Eds.) *Systemics of Emergence: Research and Development* Boston: Springer (pp. 715–722).

Porter, T. (1994). Categorical shape theory as a formal language for pattern recognition? *Ann. Math. Artif. Intell., 10*, 25–54.

Porter, T. (2002). What 'shape' is space-time? Preprint arXiv:gr-qc/0210075.

Pratt, V. (1995). Chu spaces and their interpretation as concurrent objects. *Lect. Notes in Comp. Sci., 1000* (1995), 392–405.

Pratt, V. (1999). Chu spaces. *School on Category Theory and Applications (Coimbra 1999)*, Vol. 21 of *Textos Mat. Sér. B*, University of Coimbra, Coimbra. (pp. 39–100).

Pratt, V. (1999). Chu spaces from the representational viewpoint. *Ann. Pure Appl. Logic, 96*, 319–333.

Pratt, V. (2000). Higher dimensional automata revisited. *Math. Struct. Comp. Sci., 10*, 525–548.

Raptis, I. & Zapatrin, R. R. (2001). Algebraic description of spacetime foam. *Class. Quant. Gravity, 18*, 4187–4212.

Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach.* Cambridge, MA: MIT Press.

Rosen, R. (1986). *Theoretical Biology and Complexity.* New York: Academic Press.

Rosenblatt, F. (1961). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Washington DC: Spartan Books.

Rummelhart, D. E., Smolensky, P., McClelland, J. L. & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In: McClelland, J. L. Rummelhart, D. E. & The PDP Research Group (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 2.* Cambridge MA: MIT Press (pp. 7–57).

Sakahara, K. & Sato, T. (2008). Construction of preference and information flow: I. COE Discussion paper F-215, 2008. Tokyo: Graduate School of Economics, University of Tokyo. (http://www.e.u-tokyo.ac.jp/cemano/research/DP/documents/coe-f-215.pdf)

Sakahara, K. & Sato, T. (2011). On understanding experiences of disability. In: Matsui, A. et al. (Eds.) *Creating a Society for All: Disability and Economy* Leeds, UK: The Disability Press (pp. 98–108).

Schorlemmer, M. (2002). Duality in knowledge sharing. Informatics Research Report EDI-INF-RR-0134. Division of Informatics, University of Edinburgh.

Schorlemmer, M. & Kalfoglou, Y. (2005). Progressive ontology alignment for meaning coordination: An information-theoretic foundation. Proceedings of the 4th Internationl Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2005). Utrecht, Netherlands.

Scott, D. S. (1982). Domains for denotational semantics. *Lecture Notes in Computer Science, Vol. 140* Berlin: Springer (pp. 577–613).

Seely, R. A. G. (1989). Linear logic, *-autonomous categories and cofree algebras. In: Gray, J. & Scredov, A. (Eds.) *Categories in Computer Science and Logic* (Contemporary Mathematics, 92) American Mathematical Society (pp. 577–613).

Seligman, J. (2009). Channels: From logic to probability. In: Sommaruga, G. (Ed.) Formal Theories of Information. (Lecture Notes in Computer Science, 5363) Berlin: Springer (pp. 193–233).

Shanahan, M. (2012). The brain's connective core and its role in animal cognition. *Phil. Trans. R. Soc. B, 367*, 2704–2714.

Simons, P. (1987). *Parts: A Study in Ontology.* Oxford, UK: Oxford University Press.

Smith, B. (2003). Ontology. In: Floridi, L. (Ed.) *Blackwell Guide to the Philosophy of Computing and Information.* Oxford, UK: Blackwell (pp. 155–166).

Sorkin, R. D. (1991a). Finitary substitute for continuous topology. *Int. J. Theoret. Phys., 30*, 923–947.

Sorkin, R. D. (1991b). Spacetime and causal sets. In: D'Olivo, J. C. et al. (Eds.) *Relativity and Gravitation: Classical and Quantum.* Singapore: World Scientific (pp. 150–173).

Sowa, J. F. (2006). Semantic networks. In: Nadel, L. (Ed.) *Encyclopedia of Cognitive Science.* Hoboken, NJ: Wiley.

Spanier, E. H. (1966). *Algebraic Topology.* New York: Springer.

Spivak, D. I. & Kent, R. E. (2012). Ologs: A categorical framework for knowledge representation. *PLoS ONE, 7*, e24274.

van Benthem, J. (2000). Information transfer across Chu spaces. *Logic J. IGPL, 8*, 719–731.

Wallace, R. (2005). *Consciousness: a Mathematical Treatment of the Global Neuronal Workspace.* New York: Springer.

Zhang, G.-Q. & Shen, G. (2006). Approximable concepts, Chu spaces, and information systems. *Theory Appl. Categories, 17*, 80–102.

Zimmer, H. D. & Ecker, U. K. D. (2010). Remembering perceptual features unequally bound in object and episodic tokens: Neural mechanisms and their electrophysiological correlates. *Neurosci. Biobehav. Rev., 34*, 1066–1079.