Preprint of a paper in press at *J. Exper. Theor. Artif. Intell.*, https://doi.org/10.1080/0952813X.2020.1836034

# Information flow in context-dependent hierarchical Bayesian inference

Chris Fields
23 Rue des Lavandières
11160 Caunes Minervois, FRANCE
fieldsres@gmail.com
ORCID: 0000-0002-4812-0744

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

October 12, 2020

**Abstract**

Recent theories developing broad notions of context and its effects on inference are becoming increasingly important in fields as diverse as cognitive psychology, information science and quantum information theory and computing. Here we introduce a novel and general approach to the characterization of contextuality using the techniques of Chu spaces and Channel Theory viewed as general theories of information flow. This involves introducing three essential components into the formulism: events, conditions and measurement systems. Incorporating these factors in relationship to conditional probabilities leads to information flows both in the setting of Chu spaces and Channel Theory. The latter provides a representation of semantic content using local logics from which conditionals can be derived. We employ these features to construct cone-cocone diagrams, commutativity of which enforces inferential coherence. With these we build a scale-free architecture incorporating a Bayesian-like hierarchical structure, in which there is an interpretation of active inference and Markov blankets. We compare this architecture with other theories of contextuality which we briefly review. We also show that this development of ideas conveniently accommodates negative probabilities, leading to the notion of signed information flow, and address how quantum contextuality can be interpreted within this model. Finally, we relate contextuality to the Frame Problem, another way of characterizing a fundamental limitation on the observational and inferential capabilities of finite agents.

**Keywords**: Chu space, Channel Theory, Bayesian inference, Contextuality, Information flow, Local logic, Cone-Cocone Diagram, Active Inference, Markov blanket, Frame Problem.

# Contents

**1 Introduction** **3**

**2 A background to Chu spaces and examples** **5**
2.1 Chu spaces and Chu morphisms . . . 6
2.2 Examples . . . 7

**3 A Chu space formulism of context** **9**

**4 A brief review of related probabilistic models for contextuality** **11**
4.1 The basic ingredients of CbD . . . 11
4.2 The contextual fraction . . . 12

**5 Interpretative issues: Preparation, measurement, and (quasi) probabilities** **13**
5.1 Measurement "conditions" . . . 13
5.2 Observer-dependent probabilities . . . 14
5.3 Signed probability measures . . . 15

**6 Basic concepts of Channel Theory** **16**
6.1 Classifications and Infomorphisms . . . 16
6.2 Information Channels of classifiers . . . 18
6.3 Sequents and conditional probabilities . . . 19
6.4 Constructing Cone-Cocone Diagrams . . . 20

**7 Context-compliant Hierarchical Bayesian networks** **21**
7.1 Channel Theoretic contextuality . . . 21
7.2 Commutativity requires co-deployable observables . . . 22
7.3 A Channel Theoretic interpretation of intrinsic contextuality . . . 22
7.4 A signed CCCD . . . 26

**8 Context-compliant Hierarchical Bayesian inference and contexts in practice** **26**
8.1 The Bayesian picture . . . 26
8.2 Contexts in practice I: The idea of active inference . . . 28
8.3 Contexts in practice II: The Frame Problem . . . 28

**9 Conclusion** **30**

**A Appendix: Theories and Local Logics in Channel Theory** **31**

# 1 Introduction

The observable behavior of biological, psychological, and social systems characteristically depends on the context of observation in ways other than those expected on the basis of prior probabilities or anticipated in an experimental design. Indeed, to ask how can we tell by observation the "true" context of an observed behavior is to ask a fundamental question in the psychology of perception. Such "contextuality" in the behavior of complex systems has traditionally been regarded as a practical problem of limited experimental control that could, in principle, be eliminated by obtaining more uniform experimental subjects and achieving better control of experimental conditions. This traditional view of contextuality as merely a practical limitation is, however, increasingly under strain. Over 50 years ago, Bell (1966) and Kochen and Specker (1967) proved that quantum theory, the foundation not only of modern physics but of much current technology, requires a form of contextuality even for elementary particles (see Mermin, 1993; Abramsky and Brandenburger, 2011; Dzhafarov and Kon, 2018; Fine, 1982; Khrennikov, 2009, for reviews and results). Briefly, values cannot be assigned to observable properties of elementary particles without specifying which properties are to be simultaneously observed. Here the "problem" of context dependence is not practical, but rather a strict consequence of the mathematical formulism of quantum theory. Numerous experiments have now confirmed that the observable behavior of elementary particles is, indeed, context-dependent as quantum theory requires (e.g. Bartosik et al., 2009; Kirchmair et al., 2009). If contextuality is *intrinsic* to elementary particles – or to the process of measuring properties of elementary particles, however executed or controlled – could it not also be the case that the observable behavior of biological, psychological, and social systems is intrinsically contextual, and that no amount of experimental and theoretical improvement can render such behavior context-free?

Determining whether a particular instance of observed behavior of a complex system, particularly one with memory, exhibits intrinsic contextuality is not straightforward. When a "context" is superficially ascribed to a behavior, many factors can contribute, e.g. the type of questions asked, the responses to these questions, the ambient environment of the study, the type of sample space, etc. In quantum systems, a set of measurements (i.e. a set of probability distributions of measurement outcomes) exhibits "contextuality" (i.e. exhibits "intrinsic" or "true" contextuality) if it cannot be characterized by a mathematically-consistent, context-free, joint probability distribution. Asking whether a set of measurement outcomes exhibits intrinsic contextuality is, in this case, asking whether a consistent (globally defined, globally connectable) joint probability distribution exists or not (cf. Abramsky and Brandenburger, 2011; Fine, 1982). Various formal methods for answering this question in a quantum-theoretic setting in which "no-signalling" conditions preventing classical communication can be imposed have been developed (see e.g. Abramsky and Brandenburger, 2011, 2014; Abramsky, Barbosa and Mansfield, 2017; Barbosa et al., 2019; Döring and Frembs, 2019; Frembs and Döring, 2019); cf. Kochen and Specker (1967). The formulism of *Contextuality-by-Default* (CbD) (Dzhafarov and Kujala, 2017a; Dzhafarov, Cervantes and Kujala, 2017b; Dzhafarov and Kon, 2018) approaches such questions by labeling contexts, and including them explicitly in conditional probabilities, rendering the probability distributions context-dependent. A set of random variables within a set of contexts exhibits intrinsic contextuality, in the latter formulism, whenever distinct contexts induce a difference between the distributions of random variables that exceeds the difference between the individual, within-context distributions of those variables. This criterion effectively generalises the intrinsic contextuality of quantum theory to the case in which other properties of a context, e.g. the order in which questions are asked, also affect the distribution of a variable of interest. This kind of "direct influence" is generally avoided by experimental design, particularly by imposing no-signalling conditions, in physics but is virtually inescapable when studying complex, macroscopic systems. Dzhafarov, Kujala and Cervantes (2016b) showed that a variety of "context effects" previously reported in the psychological literature could in fact be explained by such direct influences, and hence were not cases of intrinsic contextuality. Cervantes and Dhzafarov (2018) and Basieva et al. (2019) employed CbD to design psychological experiments that do demonstrate intrinsic contextuality, showing unambiguously that this principled form of contextuality occurs in complex, macroscopic systems as well as in simple systems at the microscale.

While these various approaches to contextuality to date have focused on its application to *observed* phenomena, the possibility of intrinsic contextuality in biological, psychological and social systems also raises immediate issues for the theoretical modeling of *observers*, particularly human observers. All predictive-coding or "Bayesian brain" approaches, in particular, require the observer to implement well-defined prior probability distributions (e.g. Knill and Pouget, 2004; Bubic, von Cramon and Schubotz, 2010; Clark, 2013) or sampling methods that are well-behaved in relevant limits (Sanborn and Chater, 2016). The concept of variational free energy on which active inference models are based (Friston, 2010; Friston, Kilner and Harrison, 2006; Friston et al., 2015b) becomes undefined if self-consistent prior probability distributions cannot be constructed. As such models have also been shown to apply to general biological systems (Friston, 2013; Friston et al., 2015a; Kuchling et al., 2019), this latter concern applies to models at multiple scales.

Our goals in this paper are two-fold: first, to model intrinsic or "true" contextuality using the general, category-theoretic methods of Chu spaces and Channel Theory, and second, to employ this formulation to reconstruct hierarchical Bayesian inference in a context-dependent way. We have previously shown how inference can be represented using the formulism of this paper, employing only the quantum theory of separable systems and the thermodynamics of measurement interactions (Fields and Glazebrook, 2020a). Here we begin by reviewing the category-theoretic formulism of Chu spaces and morphisms, with examples in §2. We use the tools from §2 in §3 to develop an initial category-theoretic formulation of contextuality that allows us to distinguish the complementary roles played by "preparation procedures" and "measurement contexts" in general observational settings. We briefly review some other formal models of contextuality, with which ours is consistent, in §4, and further reviewing some interpretational issues around the preparation-measurement distinction in §5.1 and §5.2; these become important in §7. In §5.3 we briefly consider a formal extension to signed probability measures, where "signed partitions" and "signed events" can be introduced (cf. Abramsky and Brandenburger, 2014). We then turn in §6 to the semantically-richer methods of Channel Theory, which provide a more explicit representation of inferential structure in the form of sequents and (regular) theories that give rise to local logics and maps, called *(logic) infomorphisms*, between them (Barwise and Seligman, 1997; Seligman, 2009). Appendix §A provides a specified explanation of these terms along with their formal definitions. It is from the sequents of these local logics that we are able to derive "conditionals" (in §6.3) such that the information channels function as inferential mechanisms. This construction is made explicit when we introduce *Cone-Cocone Diagrams* (CCCDs; Fields and Glazebrook, 2019a) as a natural, scale-free representation of Bayesian inference, and then integrate "contextual" elements into this representation. Using the CCCD representation, we are able in §7 to construct arbitrary (quasi-)hierarchical Bayesian networks, as required for active inference formulated in terms of *co-deployable observables* (Definition 7.1). We state and prove, in §7.3, our central result that failure of diagram-level commutativity in a CCCD corresponds to inadmissibility of a self-consistent probability distribution, and hence to *intrinsic contextuality*, as we define it in this paper, as a condition of non-co-deployability of measurement operators (again see Definition 7.1). Thereafter we apply this development of ideas to recover, among other things, a new theoretical and inferential role for the Markov blanket concept (Clark, 2017; Friston et al., 2015a; Kuchling et al., 2019): it is both the "locus" at which free energy is minimized and the epistemic barrier that keeps the contextual information hidden from the observer. Significantly, this holds true for *any* physical system interacting with its environment in which information is preserved, be it classical or quantum (Fields and Marcianò, 2019a). We also use these ideas to formulate a precise and strong version of the Frame Problem (McCarthy and Hayes, 1969, see Shanahan (2016) for review), and to show how this classic problem from AI relates to contextuality.

While the results presented here are primarily technical, they open a path to applications in a number of areas where contextuality has largely been neglected, but is important in practice. We have, in particular, previously employed the CCCD representation used here to model cognitive studies of visual object identification and categorization, including mereological (i.e. part-whole) classification (Fields and Glazebrook, 2019b). As we discuss below (§7.3) in connection with the "Snow Queen" experiment of Cervantes and Dhzafarov (2018), context effects on categorization can manifest as semantic inconsistencies that render choice behavior incoherent with reported beliefs. We have also shown how to reconstruct dual process theories of cognition (e.g. Evans, 2008) within the CCCD framework (Fields and Glazebrook, 2020b). This latter work develops a representation of hierarchical Bayesian inference within a Global Neuronal Workspace (GNW) architecture (Dehaene and Naccache, 2001; Shanahan and Baars, 2005); however, it does not consider contextuality. The current construction renders this previous model contextuality-compliant, and provides the building blocks needed to incorporate context switching as well as attention switching into GNW models. Our main formal result, Theorem 7.1 shows, consistent with the fundamental information-physics considerations discussed in Fields and Glazebrook (2020a), that intrinsic contextuality can always be associated with non-commutativity, and hence significant, non-direct (in the sense of Dzhafarov, Kujala and Cervantes, 2016b) order effects, between observations and/or actions. This provides an operational criterion for identifying application settings in which intrinsic contextuality can be expected; we discuss several examples in §7.3. Naive "sleeping dog" heuristics can be expected to fail as Frame Problem solutions in such settings. Finally, and from a more probability-theoretic point of view, our results further support those of Dzhafarov and Kon (2018) in showing that the distinction between "quantum" and classical probabilities lies not in any ontological difference, but rather in what has been explicitly labeled.

## 2 A background to Chu spaces and examples

For the reader's benefit, we start with a short primer on the very basic concepts of Category Theory. This is a general mathematical language for describing objects and relations (see e.g. Adámek, Herrlich and Strecker, 2004; Awodey, 2010). A *category* $\mathfrak{C}$ consists of collections of *objects*, and *arrows* (i.e. directed relations, or morphisms) between objects, satisfying two requirements: 1) arrows compose associatively, i.e. for objects $A, B, C, D$, if $f : A \to B,\ g : B \to C$, and $h : C \to D$, then $hgf : A \to D$, and 2) each object has an identity arrow $\mathrm{id}_A : A \to A$. Mathematically paired concepts such as sets with functions, sets with relations, vector spaces with linear mappings, groups with group homomorphisms, topological spaces with continuous mappings, measure spaces with measurable functions, as the respective objects and arrows, are familiar examples of categories.

## 2.1 Chu spaces and Chu morphisms

One of the simplest categories is the category of Chu spaces (Barr, 1979, 1991; Pratt, 1995, 1999a), any object in which is defined as follows:

**Definition 2.1.** Let $K$ be a set, with no structure assumed. A *Chu space* $\mathcal{A} = (A, r, X)$ *over* $K$ consists of sets $A$ and $X$, and a *satisfaction relation* (or *evaluation function*) $r : A \times X \longrightarrow K$.

Often $A$ and $X$ are referred to as sets of "objects" and "attributes", respectively.

**Example 2.1. A basic Chu space**

Chu spaces frequently give rise to matrix representations. As a simple example, take a set (of objects) $A = \{a, b, c\}$, with $K = \mathbf{2} = \{0, 1\}$. This can be represented as the Chu space:

| $r$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 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 |

It is often the case that Chu spaces may be *separable*, *extensional* or *biextensional* (i.e. both separable and extensional) or admit a *biextensional collapse* (see e.g. Pratt, 1995, 1999a; Zhang and Shen, 2006, for precise details). These terms can be conveniently explained thanks to the matrix representation: '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 and columns of attributes are factored out. In practice this removes unnecessary repetitions in the content of information, and hence minimizes the amount of processing required by a given algorithm.

By associating objects with attributes, Chu spaces provide a natural model of the process of categorizing objects by their attributes; the categorization process implements the satisfaction relation $r$. They easily generalize to multi-valued satisfaction relations, e.g. relations satisfied with some probability. The arrows in the category of Chu spaces relate one categorization process to another:

**Definition 2.2.** A *Chu transform (morphism)* of a Chu space $\mathcal{A} = (A, r, X)$ to a Chu space $\mathcal{B} = (B, s, Y)$, is a pair of functions $(f, \bar{f})$ with $f : A \longrightarrow B$, and $\bar{f} : Y \longrightarrow X$, such that for all $a \in A$, and $y \in Y$, we have $s(f(a), y)$ if and only if $r(a, \bar{f}(y))$.

The above definition specifies an adjointness or continuity principle for Chu spaces. These transforms constrain both the objects and attributes of the Chu spaces to which they relate, forcing the two categorization processes to "line up" in the way one would intuitively expect. If we now consider Chu transforms $(f, \bar{f}) : \mathcal{A} \to \mathcal{B}$ and $(g, \bar{g}) : \mathcal{B} \to \mathcal{C}$, where $\mathcal{C} = (C, t, Z)$ is a third Chu space, the Chu transform composition is given by $(g, \bar{g})(f, \bar{f}) = (gf, \bar{f}\bar{g}) : \mathcal{A} \to \mathcal{C}$, which is a Chu transform because $t(gf(a), z) = s(f(a), \bar{g}(z)) = r(a, \bar{f}\bar{g}(z))$. Chu spaces valued in $K$ along with their Chu transforms as composed, give rise to a category denoted by $\mathbf{Chu}(\mathbf{Set}, K)$.

Chu spaces are more general than topological spaces, and they have been extensively applied in theoretical computer science and event process theories (Pratt (1995, 1999a,b, 1997); see also Barwise and Seligman (1997)), and in physical and observational models (Abramsky, 2012; Fields and Glazebrook, 2020a; Gratus and Porter, 2006). In working set-theoretically, there is considerable scope for the choice of arguments in the Chu space, as well as in the choice of the corresponding relation $r$. These include probabilistic (in particular, conditional) relations (Allwein, Moskowitz and Chang, 2004; Allwein, Yang and Harrison, 2011; Kreinovich, Liu and Nguyen, 1999; Nhuy and Van Quang, 2001; Pratt, 1995), fuzzy-type relations (Nguyen et al., 2001; Papadopoulos and Syropoulos, 2000), spatial observations (Gratus and Porter, 2006), object identification and mereological reasoning (Fields and Glazebrook, 2019b), and dual process theories of cognition (Fields and Glazebrook, 2020b); numerous examples are discussed in Fields and Glazebrook (2019a). A significant fact is that every category $\mathfrak{C}$ whose arrows form a set $K$ (i.e. every 'small' category), embeds fully into $\mathbf{Chu}(\mathbf{Set}, K)$, as shown in Pratt (1996, 1997).* Indeed all theories of 'relational structures' with and without topological structure can be modeled by Chu spaces (Pratt, 1997); examples include studies pertaining to analogy and metaphor (e.g. Brown and Porter, 2006; Fields, 2011, 2013a; Gentner, 1983; Gentner and Markman, 1997; Old and Priss, 2001). Below we list several such examples with some basic details included.

## 2.2 Examples

**Example 2.2. Topological spaces**

Consider a topological space $(A, \mathcal{U})$, where $A$ is the set of points, and $X = \mathcal{U}$ is the set of open sets. Taking $r = \in$ (membership), and $K = \mathbf{2} = \{0, 1\}$, gives a Chu space $(A, r, X)$. In the matrix representation, the space is extensional (i.e. 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: $r(a, U) = 1$ implies $a \in U$, and $r(a, U) = 0$ implies $a \notin U$, given an open set $U$. In this way the category **Top** of topological spaces along with their continuous maps, fully embeds as $\mathbf{Top} \longrightarrow \mathbf{Chu}(\mathbf{Set}, \mathbf{2})$. For further specifics in the topological context, see Pratt (1999a).

**Example 2.3. Types as processes**

In Pratt (1997), Example 2.2 generalizes to a *type* when relaxing the K-valued membership relation to $r(a, x)$ indicating "the degree to which" the point $a$ belongs to the open set $U$ (cf. Nguyen et al., 2001; Papadopoulos and Syropoulos, 2000). A *process* is specified when taking $K$ to be a set of "atomic" states, $A$ a set of 'events', $X$ a set of "global" states. The process is then a sequence of events, each of which projects some atomic state out of the global state. In this case $r(a, x)$ is the atomic state picked out by an event $a$ occurring in a global state $x$. Taking $a_t \in A$ to be a measurement made at time $t$ using some operator $M$, and $X$ to be the set of states of some quantum system $S$, $r(a_t, x)$ becomes the observational outcome obtained by acting on $S$ with $M$ at $t$.

**Example 2.4. Probability spaces**

*Basically this means that the induced mapping $F : \mathfrak{C} \longrightarrow \mathbf{Chu}(\mathbf{Set}, K)$ realizes $F(\mathfrak{C})$ as an embedded subcategory of $\mathbf{Chu}(\mathbf{Set}, K)$ which consists of some objects of the latter and all of the arrows between them (see e.g. Adámek, Herrlich and Strecker, 2004; Awodey, 2010). An example is the category of conditional probabilities (Culbertson and Sturtz, 2013) whose objects consist of countably generated measurable spaces, and whose arrows belong to a semigroup of Markov kernels.

Recall that a probability space $\mathcal{P}$ consists of a triple $\mathcal{P} = (A, \Sigma, \mu)$ where $A$ is a set of events, $\Sigma$ is a $\sigma$-algebra over on $A$, and $\mu$ is a probability measure representing the probability that a given event occurs. We define a Chu space $(A, r, \Sigma)$, where $r(a, z)$ if and only if $a \in z$. This is realized by a probability in the usual way via $r : A \times \Sigma \longrightarrow [0, 1]$ (see e.g. Allwein, Yang and Harrison, 2011; Allwein, Moskowitz and Chang, 2004; Kreinovich, Liu and Nguyen, 1999; Seligman, 2009). Particularly relevant to the inferential processes studied here, there are several ways to define conditional probabilities via Chu spaces (or Classifications – see below) as demonstrated in Allwein, Yang and Harrison (2011); Allwein, Moskowitz and Chang (2004); Kreinovich, Liu and Nguyen (1999); Nguyen et al. (2001); Nhuy and Van Quang (2001). The most direct way is to assume that we have, in the usual probabilistic sense, two sets of events $A$ and $X$ that are subsets of the same probability space $\mathcal{P}$, and a function $r : A \times X \longrightarrow [0, 1]$, taken to be $r(a, x) = p(a|x)$ when the latter is defined, to construct the Chu space $(A, r, X)$ (see e.g. Kreinovich, Liu and Nguyen, 1999).

**Example 2.5. Bayesian belief network**

Following Schorlemmer (2002), let $V = \{V_1, \ldots, V_n\}$, $n \geq 1$, be a set of variables forming a configuration $b_V$, the latter defined as a conjunction of truth-valued assignments to the variables in $V$. Let $\mathcal{B}_V = \{b_V\}$ denote the set of all such configurations. A Bayesian belief network $\mathfrak{B} = (\mathfrak{G}, \mathfrak{P})$ consists of i) a directed acyclic graph $\mathfrak{G} = (V, \mathsf{Arr})$ consisting of nodes $V$ and a set of arrows $\mathsf{Arr}$, and ii) a set $\mathfrak{P} = \{P_{V_i} : V_i \in V\}$ of real valued functions $P_{V_i} : \mathcal{B}_{V_i} \times \mathcal{B}_{\mathsf{prec}(V_i)} \longrightarrow [0, 1]$, where $\mathsf{prec}(V_i)$ denotes the set of variables that immediately precede $V_i$ in the graph. The set of functions $\mathfrak{P}$ uniquely determines a joint probability function $P_V : V \longrightarrow [0, 1]$ extended to conditional probabilities using Bayes' rule, thus leading to a function $\mathsf{P} : \mathcal{B} \times \mathcal{B} \longrightarrow [0, 1]$, where $\mathcal{B}$ denotes the set of all configurations, i.e. $\mathcal{B} = \{b_W : W \subseteq V\}$. From this we obtain a Chu space corresponding to $\mathfrak{B}$, given by $(\mathcal{B}, r, \mathcal{B})$ where $r$ is a real-valued function as defined via the conditional probabilities across all configurations; i.e. $r(b_U, b_W) = \mathsf{P}(b_U|b_W)$.

**Example 2.6. Event space structure**

Let $A$ be a countable set of general events ('atomic' events, events observed or experienced, the imposition of boundaries, etc.) causally ordered by a partial ordering "$\leq$": given $a_1, a_2 \in A$, this means that $a_1$ precedes $a_2$ in time; equivalently, if $a_2$ has happened to occur, then so has $a_1$. As in Pratt (1997), $A$ can be conveniently labelled by some algebra of "actions" denoted by $\Lambda$, with assignment $\lambda : A \longrightarrow \Lambda$, such that with ordering and labelling we obtain a *partially ordered multiset* $(A, \leq, \Lambda)$. In order to deal with possibly conflicting events, we introduce a *conflict relation* $\#$, a symmetric irreflexive binary operation which specifies the non-occurrence of two events. Hence, $a_1 \# a_2$ rules out $a_1, a_2$ occurring simultaneously in a process, and a choice has to be made between them once preceding events have occurred. The axiom of conflict is that if $a_1 \# a_2$ and $a_2 \leq a_3$, then $a_1 \# a_3$. Following Pratt (1997), we thus arrive at the Chu space of events $(A, r, Z)$, where $Z$ denotes the set of order ideals of $(A, \leq)$ that do not contain both $a_1$ and $a_2$ when $a_1 \# a_2$, and where $r : A \times Z \longrightarrow K$.[†] Other approaches to event space structures involving "probabilistic runs" of $(A, \leq, \#)$ are studied in Varacca, Völzer and Winskel (2006); Winskel (1982).

**Example 2.7. Chu flows**

[†] In Pratt (1997), $K = \mathbf{2}$ is mainly taken, but other possibilities are discussed when working with the biextensional (full subcategory) $\mathbf{chu}(\mathbf{Set}, K)$ of $\mathbf{Chu}(\mathbf{Set}, K)$.

Given Chu spaces that are related by a Chu transform, we may ask what type of information the transform preserves. A *Chu flow* of information (van Benthem, 2000, cf. Barwise and Seligman (1997)) may specified by a "flow formula" constructed from the elements of the following schema:

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

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. van Benthem (2000) show that for finitely-structured Chu spaces $A$ and $B$, the existence of a Chu transform $A \longrightarrow B$ is equivalent to every flow formula valid in $A$ being valid in $B$ as well. The transform $A \longrightarrow B$ can, in this case, be viewed as "transporting" the information encoded in valid flow formulas from $A$ to $B$; it can thus be thought of informally as a "channel" from $A$ to $B$, and as implicitly providing a sense of "spatial" and/or "temporal" separation between $A$ and $B$ (see also Krötzsch, Hitzler and Zhang, 2005). This flow of information through channels will be more fully specified in terms of "classifiers" and "infomorphisms" in §6.1.

**Example 2.8. Kullback-Leibler divergence and Free Energy**

This example provides an initial formulation of a Bayesian-like inferential mechanism using the Chu-space concepts developed above; a richer development with Channel Theory will be described in §6. Given discrete probability distributions $P$ and $Q$ on the same space $X$, the Kullback-Leibler Divergence (KLD) $D_{KL}(P \parallel Q) = \sum_x P(x)\ \log(P(x)/Q(x))$ measures the "surprise" of an observer encountering $P$ when expecting $Q$. The variational free energy (VFE) of an environment characterized by $P$ for an observer with prior probabilities $Q$ is $D_{KL}(P \parallel Q)$ minus the observer's ability to predict $P$ given $Q$, i.e. the observer's ability to update their prior *before* making the relevant observation (Friston, 2010; Friston and Stephan, 2007; Friston, 2013). The KLD is seen as a "cross entropy" term, such that free energy $F$ = surprise + cross entropy. More specifically, consider a sensory input $\tilde{s}$, a world model $m$, $\vartheta$ some unknown quantity causing $\tilde{s}$, $\mu$ a quantity depending upon internal (e.g. brain) states, and a recognition density $v(\vartheta|\mu)$. Then (Friston, 2010; Friston and Stephan, 2007; Friston, 2013):

$$F = -\ln p(\tilde{s}|m) + D(v(\vartheta|\mu)\|p(\vartheta|\tilde{s})) \tag{2.2}$$

This last expression can be reformulated in terms of expectations with respect to the recognition density $v$, as:

$$F = -\langle \ln p(\tilde{s},\vartheta|m)\rangle_v + \langle \ln v(\vartheta|\mu)\rangle_v \tag{2.3}$$

(as noted in Collel and Fauquet (2015)). Minimizing free energy involves minimizing surprise and consequently, prediction error. Cross entropy is minimized by inferential updating of internal representations in tandem with revising the recognition density $v$ to a finer approximation of the "true" distribution. Clearly, any such predictive ability requires observing something about the environment that would indicate a need to adjust expectations, as demonstrated in e.g. Friston and Kiebel (2009); Friston et al. (2015b,a). This naturally points to "contextual" factors, as will be discussed throughout the remainder of this paper (see in particular §8.2 and §8.3).

## 3 A Chu space formulism of context

Here we proceed to formulate a Chu space model for dealing with contextuality, and work at the purely set theoretic level. We briefly review some related approaches to contexuality in §4. Consider then countable (in practice, finite) sets:

i) $A$ an "event" set (e.g. as in §2.6) of observed value combinations, as related to

ii) $B$ a set of conditions specifying "objects/contents" or "influences," and

iii) $R$ a set of contexts (or, in certain instances, a set of "detectors" or "methods").

We can interpret the set $B$ as $B = B^M \cup B^C$ (disjoint union), where $B^M$ contains "objects/contents" or "degrees of freedom" that are observed or measured in some event $a \in A$, and $B^C$ contains what is *not* observed in the events in $A$. From a traditional perspective, in which all observations are assumed *a priori* to be context-independent, it would seem appropriate to consider the conditional $A|B^M$, i.e. that the observed events depend only on the observed conditions. This is the assumption of local, complete measurements familiar from classical physics, or equivalently, the assumption of task-environment circumscription familiar from "good old-fashioned" artificial intelligence (Dietrich and Fields, 1996). We will, however, opt to consider a 'large' space:

$$X := B \times R = (B^M \cup B^C) \times R \tag{3.1}$$

by assuming that $A, B$ and $R$ are subsets of the same (even larger) probability space $\mathcal{P}$ as in Example 2.4.[‡] Here we do not make *a priori* assumptions about the corresponding types of probability distributions (e.g. discrete versus continuous), the nature of random variables, or possible orders of "connectedness" of distributions (however, see §4 and §7 for consideration of this latter property). Note that the formulation of (3.1) is a convenient way of expressing how contexts can be paired with, or parametrized by conditions (including unobserved conditions), and vice-versa. We then formulate the Chu space $\mathcal{A} = (A, r, X)$, where the satisfaction relation is the conditional probability $r(a, x) = p(a|x) = p(a|\{b, c\})$, whenever defined, for $a \in A, b \in B$ and $c \in R$. Collectively, the above ingredients are those we take to model a system of "observables" in what follows.

Let $A_\alpha \subset A$ be a subset of "events" occurring within a corresponding subset of "contexts" $R_\alpha \subset R$. Assuming for now that the set of conditions $B$ is fixed, and setting $X_\alpha = B \times R_\alpha$, leads to a Chu space $\mathcal{A}_\alpha = (\{A_\alpha\}, r, \{X_\alpha\})$, suitably indexed by $\alpha$. These constituents of $\mathcal{A}_\alpha$ are, in all realistic cases, considered to be finite, but we can also consider the limit in which they are countable. In principle, running through a sequence of the $\mathcal{A}_\alpha$ across some range of $\alpha$, leads to a Chu flow as in Example 2.7. This will be specified in terms of an information flow of "classifiers" in §7.1 below. Note that this flow does not in itself create couplings via marginal distributions, for instance, but it does preserve logical cohesion in information flow as explained in Example 2.7. At this stage, one may ask if imposing a joint distribution on $\{A_\alpha\}$ may be related to the existence of a *cocone (colimit)* of a system of Chu flows interpreted inferentially. To further specify this question, and to respond to it, we will use the Channel Theory formulation in §7.1, as it provides a natural representation of inference. In this richer representation, we can use the existence of a cocone to distinguish between intrinsic contextuality and a context-independent description.

We point out that this interpretation of contextuality in terms of Chu spaces (to be made more specific in §7.1) makes no "physical" or otherwise scale-dependent assumptions[§]; hence it can be expected to apply to "observations" at any scale. This immediately suggests the possibility of a hierarchical theory in which observations are made simultaneously at multiple scales, with

[‡]In physics parlance, this amounts to "going to the Church of the Larger Hilbert Space" where there are enough degrees of freedom to define pure states (cf. Chiribella, D'Ariano and Perinotti, 2016).

[§]We interpret "scale" here broadly to include not only physical (e.g. length or time) scales but also organizational scale within a complex system.

information transfer between scales contributing to the updating of prior probabilities on each measurement cycle. Notably, this a core idea of "Bayesian brain" approaches to cognition (see e.g. Friston (2010); Friston and Kiebel (2009)). In §6, while essentially continuing to work in the Chu category, we will re-formulate the model of "observation" as a flow of information through a (quasi-)hierarchical inferential process in the setting of classifications and local logics, where general countability of sets is admissible. This will accompany an alternative means of formulating conditional probabilities. First, however, we consider some issues of interpretation motivated by the original derivations of intrinsic contextuality to explain empirical phenomena in physics. We will return to these issues with the full conceptual framework of active inference in §8.2.

# 4 A brief review of related probabilistic models for contextuality

## 4.1 The basic ingredients of CbD

As mentioned in the Introduction, CbD adopts the view that *all* measurements are carried out in some non-trivial context, and hence that *all* measurement outcomes should carry a context label, with outcomes obtained in different contexts amenable to a coupling of probability distributions. For an outline of basic ideas, it will be sufficient to discuss cyclic and binary systems following Dzhafarov and Kujala (2017a); Dzhafarov, Cervantes and Kujala (2017b). One commences with a measurement system $\mathfrak{R}$ in which a typical measurement outcome, denoted $R^c_q$, is taken to be a random variable, described as follows. Let $Q$ be a set of properties representing, for instance, "objects" or "inputs" depending upon the observation or application in question; thus $q \in Q$ is called the *content* of $R^c_q$. The superscript $c$ specifies how, e.g. with what tools or instruments, $q$ is measured, and is called the *context* of $R^c_q$. In this way, a content-context (c-c) pair $(q, c)$ uniquely specifies the random variable $R^c_q$ within the system $\mathfrak{R}$ in question. It is assumed that $R^c_q$ is characterized by a distribution, based on sufficient sampling. If multiple random variables are measured in a single context (e.g. a pair $(R^c_1, R^c_2)$) these have a joint distribution. There is, however, no guarantee that random variables measured in different contexts (e.g. a pair $(R^1_2, R^2_1)$) can be stochastically related; *a priori* they are *stochastically unrelated*, and thus generally not considered to comprise well-defined "events" such as $[R^1_2 = x, R^2_1 = y]$. The formal question of interest is, then, to characterize the conditions under which a well-defined (in the sense of satisfying the Kolomogorov axioms) cross-context joint probability distribution exists.

This starts by seeing two content-sharing random variables $(R^1_q, R^2_q)$ as comprising a *connection* between contexts. In a connection, the content elements are always, by default, considered to be pairwise stochastically unrelated, and hence cannot be considered as the same random variable even if they are identically distributed. A system $\mathfrak{R}$ is said to be *consistently connected* if for every content $q$, and contexts $c, c'$ containing $q$, the distributions of $R^c_q$ and $R^{c'}_q$ are the same. Essential to this formulism is the general notion of coupling of random variables (see e.g. Lindvall (1992)). Context-induced changes in the individual distributions, i.e. how $R^1_q$ differs from $R^2_q$, are characterized by determining whether they have a *maximal coupling*.

The first step is to replace $R^1_q$ and $R^2_q$ with jointly distributed random variables $T^1_q$ and $T^2_q$ that have the same respective individual distributions. One then searches for such a replacement with the maximal value of $\mathsf{P}[T^1_q = T^2_q]$, which given a maximal coupling always exists, and is unique. If $R^1_q$ and $R^2_q$ are consistently connected, and this preserved under the maximal coupling, then $R^1_q$ and $R^2_q$ become stochastically related via introduction of the former. So as not to lose the joint distribution of $(R_q^1, R_{q'}^1, \ldots)$, and the joint distribution of $(R_q^2, R_{q'}^2, \ldots)$, a coupling of all random variables is to be taken. If, however, $R_q^1$ and $R_q^2$ are not consistently connected, they remain stochastically unrelated. In this case $\mathfrak{R}$ is deemed to be a *contextual* measurement system, where in this case "contextual" indicates intrinsic contextuality, i.e. lack of consistent cross-context joint distributions. Otherwise, $\mathfrak{R}$ is *non-contextual*; equivalently, there exists a coupling of all random variables such that the subcouplings corresponding to connections that are maximal. In this latter case, all apparent "context effects" are due to direct influences, i.e. changes in the within-context distribution(s) of one or more random variables induced by changing the measurement context, as shown for several examples by Dzhafarov, Kujala and Cervantes (2016b).

More generally, suppose $R_q^1, \ldots, R_q^k$ $(k > 1)$ is a connection of a system. A coupling $(T_q^1, \ldots, T_q^k)$ of $R_q^1, \ldots, R_q^k$ is said to be *multimaximal* if for any $m > 1$, and for any subset $(T_q^{i_1}, \ldots, T_q^{i_m})$ of $(T_q^1, \ldots, T_q^k)$, the value of $\mathsf{P}[T_q^{i_1}, \ldots, T_q^{i_m}]$ is the largest possible among all possible couplings of $(R_q^{i_1}, \ldots, R_q^{i_m})$. In fact, there always exists a maximal coupling of any set of random variables, and a multimaximal coupling $(T_q^1, \ldots, T_q^k)$ always exists for a connection $R_q^1, \ldots, R_q^k$ with binary random variables (Dzhafarov and Kujala, 2017a, Coroll. 1). Further, a coupling of a (c-c) system is said to be *multimaximally connected* if every subcoupling of this coupling corresponding to a consistent connection of the system is a multimaximal coupling of this connection. This leads to a (c-c) system of binary random variables decreed to be *noncontextual* if it has a multimaximally connected coupling. Otherwise it is said to be *contextual*. A measure of contextuality, in terms of 'quasi-couplings', is prescribed by Dzhafarov and Kujala (2017a, Theorem 4) to which we refer. Such measures indicate the extent to which context effects *cannot* be attributed to direct influences, and hence indicate "intrinsic" or "true" contextuality, also without scale-dependent assumptions. Examples of applications of contextuality measures to demonstrate intrinsic contextuality in psychological data may be seen in Cervantes and Dhzafarov (2018) and Basieva et al. (2019).

**Example 4.1. Measuring conditional probabilities across contexts**

Let us exemplify the formulism of §3 as it relates to the above setting. Suppose that the set $X = B \times R$ in §3 comprises a set of binary random variables governed by a measurement (c-c) system as described above that is not consistently connected, and in which some subsets $\{a_i\}$ of events in $A$, and $\{b_j\}$ of conditions in $B$ are measured in every context in $R$ (Note that with this assumption, we have effectively identified $B$ with the set $Q$ of contents, and $R$ with the set $C$ of contexts.) Then there is at least one pair $c$, $c'$ of contexts, $c \neq c'$, and at least one event $a \in \{a_i\}$, and condition $b \in \{b_j\}$, such that $p(a|\{b, c\}) \neq p(a, |\{b, c'\})$. This statement has a version in a channel-theoretic representation of contextuality, which we prove below as Corollary 7.1.

## 4.2 The contextual fraction

An alternative approach to measuring intrinsic contextuality has been proposed by Abramsky, Barbosa and Mansfield (2017) (see also Barbosa et al. (2019)). Here a "measurement scenario" is formalized as an empirical model $e$ that is specified by a probability distribution expressed as a convex combination of a non-contextual model $e^{\mathrm{NC}}$ and a "no-signalling" model $e'$, i.e. $e = \lambda e^{\mathrm{NC}} + (1-\lambda)e'$, with $\lambda \in [0, 1]$. The no-signalling condition assures full statistical independence of the random variables encompassed by $e'$ (for a general discussion, see Mermin (1993)); hence it rules out direct influences within $e'$ while allowing contextuality. Relative to some global probability distribution on the outcomes to all measurements, the maximum possible, admissible value of $\lambda$ in such a de-composition of the local model $e$ is called the *non-contextual fraction* $\mathsf{NCF}(e)$ of $e$; the contextual fraction is $\mathsf{CF}(e) := 1 - \mathsf{NCF}(e)$. The model $e$ is said to be *contextual* if the corresponding family of probability distributions in Abramsky, Barbosa and Mansfield (2017) cannot itself be obtained as the marginals of the global probability distribution on the outcomes to all measurements. This has a companion interpretation in terms of the non-existence of a global section of a sheaf of distributions defined over a "measurement cover" (Abramsky and Brandenburger, 2011) (cf. Fine (1982)). As "all measurements" would include the ancillary measurements that, in the CbD formulism, say, provide the context labels, this way of thinking about contextuality can be seen as an implicit, *post facto* approach to determining whether intrinsic contextuality obtains in a data set assumed at the outset to be context-independent. Barbosa et al. (2019) have employed this approach to study the Bell inequalities (Bell, 1964, 1966) and their associated notion of "non-local" dependencies (i.e. quantum entanglement) between random variables. A key common factor of these theories pertains to non-consistently connected/non-globally definable probability distributions, which we will discuss more generally in terms of "non-commutativity" in §7. [¶]

# 5 Interpretative issues: Preparation, measurement, and (quasi) probabilities

## 5.1 Measurement "conditions"

When we assume that the subsets $R_\alpha$ of the set of contexts $R$ have not been fully characterized by observation, or have not been noted explicitly when experimental data are recorded, or are simply not taken into account when computing probabilities of experimental outcomes, contextuality becomes "quantum" contextuality in the sense of Kochen and Specker (1967). An observer only measures the $P(A|B^M)$ distributions across contexts, and finds that they violate classical probability (e.g. the Kolmogorov axioms). Both the subsets $R_\alpha$, and the unobserved "background" conditions $B^C$ amount in this case to "hidden variables" that influence the observed outcomes while remaining undetected and possibly undetectable except via the *post hoc* methods of Abramsky, Barbosa and Mansfield (2017) addressed in §4.2 above.

Why should context matter? In physics, intrinsic contextuality poses the same challenges to classical thinking as are posed by violations of Bell's theorem: either strict locality or counterfactual definiteness are ruled out (Bell, 1964, for review, again see Mermin (1993); Khrennikov (2009)). Hence intrinsic contextuality is, from this point of view, very surprising. "Direct influences" of context are not: all experiments involve a preparation step, the various aspects of which are all observed, or at least could be observed, and the outcomes of these observations – i.e. the experimental preparation itself – are obviously *intended* to influence the "experimental" outcomes that follow. It is generally assumed when discussing physics experiments that *all* relevant aspects of the preparation procedures have been observed, and hence that no unexpected direct influences will appear as unanticipated "context effects." This completeness assumption is often wrong in practice; for a spectacular case, see §6.1 of OPERA Collaboration (2012). When discussing observations of biological, psychological, and social systems, however, this completeness assumption is

[¶] The quantum-theoretic approach to contextuality is also developed by other means in e.g. Döring and Frembs (2019); Frembs and Döring (2019); Gudder (2019). In particular, Gudder (2019) introduces the notion of a "quantum channel" formulated in purely functional-analytic terms, which differs (but is possibly relatable to) the category-theoretic based Channel Theory of information as adopted in this paper.

generally acknowledged to be unjustifiable, i.e. $B^C \neq \emptyset$ is taken for granted and "context effects" are expected, though of course unpredictable in detail.

In the framework of measurement systems, as presented in e.g. Abramsky, Barbosa and Mansfield (2017); Dzhafarov, Kujala and Cervantes (2016a); Dzhafarov, Cervantes and Kujala (2017b), a "context" is just a set of (possibly ancillary) measurements. In the analysis of Kochen and Specker (1967), it is a set of measurements made with one, specified set of detectors following one, specified preparation. Is this notion of a context adequate in practice? Could there not be "context effects" between, e.g. the first and second halves of a single "experimental run" not explainable as direct influences? Contextuality, in this case, would manifest as a within-run violation of the Leggett-Garg inequality (Emary, Lambert and Nori, 2014). Here the question of how to circumscribe the unobserved degrees of freedom $B^C$ becomes relevant (Fields, 2018). We return to this question in §8.3 below.

## 5.2 Observer-dependent probabilities

The question of whether probabilities are objective features of the world or subjective beliefs of observers has been debated, with little resolution, since the time of Laplace. These issues can be given an empirical footing by considering instead the operational question of the *observer-dependence* of probabilities as measured. This highlights the relevance of the observer's observational capabilities, and for understanding what the observer does with the observations, including how they are reported to third parties, the observer's inferential capabilities and priors.

Two issues in particular arise in physics and can be carried over into discussions of observations of complex systems. First is the question of background knowledge, including both knowledge of preparation procedures and "general" situation-nonspecific knowledge that may be relevant to "knowing what to look for" in the context of an experiment and hence to the question of defining $B^C$. If Alice and Bob define $B^C$ differently, they will have implicitly and possibly unknowingly defined their contexts differently.

**Example 5.1. Alice and Bob's question of context**

The canonical experimental test for entanglement in physics is the Bell/EPR experiment, involving two observers (Alice and Bob) and two detectors, generally Stern-Gerlach devices (Bell, 1964; Mermin, 1993). If Alice and Bob are each observing their respective Stern-Gerlach device and cannot communicate, they may fail to describe their respective contexts as including a common state preparation. In the CbD formulism, for instance, this renders their observed outcomes *prima facie* stochastically unrelated, and a maximally connected coupling on these unrelated outcomes is well-defined. If, however, they possess prior knowledge of a common preparation procedure for a Bell-EPR experiment and include this knowledge in their context descriptions, their outcomes become physically correlated and their joint probability distribution violates Bell's inequality and hence violates the Kolmogorov axioms (Bell, 1964; Mermin, 1993). They may, alternatively, discover *post facto*, once they again communicate, that their outcomes are stochastically related (indeed, supra-classically related) by discovering a maximal coupling as discussed in §4.1 and above. Putting it another way, in the CbD framework it is the case that Bell's experiment does not violate the Kolmogorov axioms once the random variables involved are defined according to the framework. In this case there is a joint distribution (or coupling) of all random variables, but the coupling cannot be chosen such that subcouplings corresponding to connections are identity couplings.[‖]

This example raises the question of whether Alice and Bob can determine, by observation and/or classical communication, that they have defined the same context. If they cannot, one is placed in the position advocated by QBism (Fuchs and Schack, 2013) in which all probability assignments are strictly perspectival (Mermin, 2019).

The second issue concerns the availability and shareability of reference frames. As discussed in Bartlett, Rudolph and Spekkens (2007), physically-implemented reference frames such as meter sticks or clocks, and by extension, all physically-implemented apparatus, encode "non-fungible information," i.e. information that can be exchanged only by exchanging a physical system, not just a bit string. If a "context" is defined by the use of a state-preparation device (e.g. a laser or a particle accelerator) and a detector, it can be shared, e.g. between distant observers, only if these physical entities can be shared. Observers themselves, however, also encode non-fungible information, which without destroying the physical integrity of the observer cannot be shared (Fields and Marcianò, 2019b). Here again, we are pushed toward a position in which outcomes and their probabilities are strictly perspectival.

### 5.3 Signed probability measures

Feynman (1987) once advocated a close connection between what was conceived as "quantum" probability theory, and the admittance of negative (quasi)probabilities into the standard theory (see also Baker (1958); Scully, Walther and Schleich (1994)). The inclusion of negative probabilities here may be approached by assuming that the sets $A, B$ and $R$ of §3 are subsets of a quasiprobability space $\widetilde{\mathcal{P}}$ for which the usual Kolmogorov axioms are relaxed to accommodate negative probabilities, and where conditionals $P(A|X) = P(A, |B \times R)$ are definable. A related approach, apart from the quantum-theory method of Wigner functions, is to consider for any set $Z$, the set $\mathcal{M}(Z)$ of *signed probability measures* (Abramsky and Brandenburger, 2014). In this framework, negative probabilities arise from standard probabilities on *signed events*, and enable an alternative, systematic treatment of "no-signalling" models. Since "events" figure in our formulism, we can conveniently adopt the notion of signed events following Abramsky and Brandenburger (2014).

Consider a set of finitely supported maps $m : Z \longrightarrow \mathbb{R}$, satisfying $\sum_{z \in Z} m(z) = 1$. Measures can be extended to subsets $W \subseteq Z$ by (finite) additivity: $m(W) = \sum_{w \in W} m(w)$. Let $\mathcal{P}(Z) \subset \mathcal{M}(Z)$ denote the subset of nonnegative real-valued measures; effectively, these become the finitely-supported probability distributions on $Z$. Given a function $f : Z \longrightarrow Y$, we define a map

$$\begin{aligned} \mathcal{M}(f) : \mathcal{M}(Z) &\longrightarrow \mathcal{M}(Y) \\ m &\mapsto [y \longrightarrow \sum_{f(z)=y} m(z)] \end{aligned} \tag{5.1}$$

that pushes forward measures on $Z$ along $f$ to measures on $Y$. In this way, $\mathcal{M}$ will always be a map between probability distributions, and so let $\mathcal{P}(f) := \mathcal{M}(f)|_{\mathcal{P}(Z)}$. These conditions can be seen to be functorial:

$$\mathcal{M}(g \circ f) = \mathcal{M}(g) \circ \mathcal{M}(f), \ \mathcal{M}(\mathbf{id}_Z) = \mathbf{id}_{\mathcal{M}(Z)} \tag{5.2}$$

and likewise for $\mathcal{P}(Z)$. Using signed measures as implementing negative probabilities, the former can be interpreted as a means of "pushing the minus sign inwards". This means that basic

[‖] We thank an anonymous reader for pointing this out.

"events" are assumed to have some additional information in the form of a sign or "probability charge". Crucially, occurrences of the same underlying event of opposite sign cancel, so that negative probabilities arise from standard probabilities on signed events. The formal description starts by assigning "$\pm$" to relevant objects: given a set $E$, the signed version $E^{\pm}$ is taken to be the disjoint union of two copies of $E$, expressed as

$$E^{\pm} := \big\{(e, \zeta) : e \in E,\ \zeta \in \{+, -\}\big\}. \tag{5.3}$$

We refer to Abramsky and Brandenburger (2014) for complete details, including probabilistic signed instruction examples.

# 6 Basic concepts of Channel Theory

We now introduce an extension of the basic Chu-space formulism that builds in sufficient semantics to enable a natural representation of inference. This allows us to define a semantically-richer notion of probability that is explicitly process-oriented. Using this notion and the category-theoretic concepts of limit and colimit (Adámek, Herrlich and Strecker, 2004; Awodey, 2010), we construct a representation of bidirectional logical constraint flow, i.e. a representation of inference that is simultaneously bottom-up and top-down; this is the CCCD mentioned earlier (Fields and Glazebrook, 2019a). We then employ this representation in §7 to reconstruct Hierarchical Bayesian inference in a contextually-compliant way.

## 6.1 Classifications and Infomorphisms

For the present purposes, the most important application of Chu spaces is in modeling semantic (Dretske, 1981) or "pragmatic" (Roederer, 2005) information flow and inference, following the methods of Barwise and Seligman (1997)(see also Barwise (1997); Seligman (2009)). The fundamental concept in this case is the idea of a "Classification", or "classifier" (as we say throughout) relating "Tokens" to the "Types" that encompass them. We follow the standard notation of Barwise and Seligman (1997):

**Definition 6.1.** A *classifier* $\mathcal{A}$ is a triple $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ where $\mathrm{Tok}(\mathcal{A})$ is a set of "tokens", $\mathrm{Typ}(\mathcal{A})$ is a set of "types", and $\Vdash_{\mathcal{A}}$ is a *classification* relation between tokens and types:

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

**Example 6.1. First order language**

A first order language $L$ is a classifier, where Tok(L) consists of a set $M$ of certain mathematical structures, and Typ(L) are sentences in $L$, and $M \Vdash \varphi$, if and only if $\varphi$ is true in the token $M$. The type set of a token $M$ is the set of all sentences of $L$ true in $M$, called the *theory* of $M$ (Barwise and Seligman, 1997, Ex 2.2, p.28; see Appendix Definition A.1 for a more general treatment of this last term).

In Channel Theory, Chu transforms become "infomorphisms" which are natural maps between classifiers (Barwise and Seligman, 1997):

**Definition 6.2.** Given two classifiers $\mathcal{A} = \langle \mathrm{Tok}(\mathcal{A}), \mathrm{Typ}(\mathcal{A}), \Vdash_{\mathcal{A}} \rangle$ and $\mathcal{B} = \langle \mathrm{Tok}(\mathcal{B}), \mathrm{Typ}(\mathcal{B}), \Vdash_{\mathcal{B}} \rangle$, an *infomorphism* $f : \mathcal{A} \to \mathcal{B}$ is a pair of maps $\overleftarrow{f} : \mathrm{Tok}(\mathcal{B}) \to \mathrm{Tok}(\mathcal{A})$ and $\overrightarrow{f} : \mathrm{Typ}(\mathcal{A}) \to \mathrm{Typ}(\mathcal{B})$ such that $\forall b \in \mathrm{Tok}(\mathcal{B})$ and $\forall a \in \mathrm{Typ}(\mathcal{A})$, $\overleftarrow{f}(b) \Vdash_{\mathcal{A}} a$ if and only if $b \Vdash_{\mathcal{B}} \overrightarrow{f}(a)$.

This last condition may be schematically represented by a commutative diagram:

$$\begin{tikzcd} \mathrm{Typ}(\mathcal{A}) \arrow[r, "\overrightarrow{f}"] \arrow[d, dash, "\Vdash_{\mathcal{A}}"'] & \mathrm{Typ}(\mathcal{B}) \arrow[d, dash, "\Vdash_{\mathcal{B}}"] \\ \mathrm{Tok}(\mathcal{A}) & \mathrm{Tok}(\mathcal{B}) \arrow[l, "\overleftarrow{f}"'] \end{tikzcd} \tag{6.2}$$

Intuitively, an infomorphism is a way of transmitting information from one classifier to another, so that, e.g. "b is type B" can encode, or represent, the information "a is type A". The relations $\Vdash_{\mathcal{A}}$ and $\Vdash_{\mathcal{B}}$ are explicitly regarded as enforcing semantic, not merely syntactic or set-theoretic constraints, rendering both classifications and infomorphisms intrinsically semantic notions (to be elaborated upon in §6.2 below). Thus, "information" here is not simply reduced to a quantity of bits (as is the case for Shannon information; see e.g. Cover and Thomas (2006)), but it is rather the set of logical constraints as imposed by Definition 6.2, and thus can be viewed as *pragmatic information* as proposed in Roederer (2005).[**]

**Example 6.2. Messages and contents**

As a second straightforward example, consider the classification $\mathbf{M} = \langle Messages, Contents, \Vdash_{\mathbf{M}} \rangle$ where Messages are classified by their Contents (Allwein, Moskowitz and Chang, 2004), and a further 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, $m^f \Vdash_{\mathbf{M}} C$ if and only if $m \Vdash_{\mathbf{M}'} C^f$.

The idea of an infomorphism as a mapping between classifiers provides the basic building block for constructing multi-level, quasi-hierarchical classification systems. Like the connections between "processing layers" in brains, infomorphisms are intrinsically bidirectional. In contrast to the superficially-similar treatment of Ehresmann and Vanbremeersch (2007); Ehresmann and Gomez-Ramirez (2015), infomorphisms here are bidirectional maps between sets of logical relations as will be specified below.

[**] Allwein (2004) makes similar remarks when pointing to Barwise and Seligman (1997): "How do remote objects, situations and events carry information about one another without any substance moving between them?". The general qualitative framework of Barwise and Seligman (1997), though not pinpointing specific measures of data flow within passive messaging, as in the Shannon information case, does nevertheless provide a flow of logical/semantic reasoning, as noted in e.g. Allo (2009), when the logical constraints of Definition 6.2 lead to structured inference. Given the premise that information is a physical mode of distinctions and the relationships between them, and causation is understood as transfer of information, Collier (2011) applies the logical formulism of Channel Theory to argue that causation itself may be viewed as a form of computation in view of the regular relations in a distributed system. In consonance, Old and Priss (2001) regard such information flow as a "conceptual" channel between source and target. Accordingly, each classifier accommodates a "context" in terms of its constituent tokens and types.

### 6.2 Information Channels of classifiers

The specifics of transmitting information in Barwise and Seligman (1997) motivates defining an *information channel* **Chan** as a finite 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**:

$$\begin{tikzcd} & \mathbf{C} & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[u, "f_2"'] & \ldots\ \mathcal{A}_k \arrow[ul, "f_k"'] \end{tikzcd} \tag{6.3}$$

The core $\mathbf{C}$ functions as a carrier of information flow between the $f_i$, and hence between the component classifiers $\mathcal{A}_i$. Intuitively, the channel can be viewed as a "wire" connecting two agents (i.e. classifiers) to a "blackboard" or other shared memory via which they can exchange information. As the shared memory $\mathbf{C}$ is itself a classifier, it admits a structure regulating how information is written to and read from it; for example, it may be a "smart blackboard" that incorporates a function such as a multi-language translation (Fields and Glazebrook, 2019a). This idea of a channel core as a shared memory is developed in the setting of interacting channels forming a *distributed system* (Barwise and Seligman, 1997), which underlies the following constructions.

**Example 6.3. A distributed system**

Following Barwise and Seligman (1997, p.89) consider two information channels sharing a common classifier $\mathcal{A}_2$ as below. Suppose the first channel leading to the classifier $\mathbf{B}_1$ represents the examination of a map, and captures the information a person obtains by doing so. The second channel leading to the classifier $\mathbf{B}_2$ represents the informational relationship between the map and the region it illustrates. These channels are clearly distinct but coupled: the information content of $\mathbf{B}_1$ is related, but far from identical, to that of $\mathbf{B}_2$. This coupling can be depicted as:

$$\begin{tikzcd} & \mathbf{B}_1 & & \mathbf{B}_2 & \\ \mathcal{A}_1 \arrow[ur, "f_1"] & & \mathcal{A}_2 \arrow[ul, "f_2"'] \arrow[ur, "f_3"] & & \mathcal{A}_3 \arrow[ul, "f_4"'] \end{tikzcd} \tag{6.4}$$

The sense in which channels encode sets of mutual constraints holding between classifiers is further elaborated in Barwise and Seligman (1997); Fields and Glazebrook (2019a) where the classifier concept is extended to that of a *local logic*, the formal details of which are outlined in the Appendix (Definition A.4). For now, we can broadly explain this as follows. Intuitively, a classifier is extended to a local logic by specifying a subset (possibly a singleton) of tokens satisfying all of the types of some given (regular) *theory*, as exemplified in Example 6.1, that specifies the logical aspects of a given situation (see Definitions A.1 and A.2). The theory is *regular* in so far that it regulates the structural properties to which the system in question adheres. Accordingly each local logic incorporates its own regular theory (Definition A.4). The governing rules are expressible in mathematical (logical) terms with respect to the information flow. Thus an infomorphism can be extended to a *logic infomorphism* that preserves this additional structure (Definition A.5). Intuitively, a local logic "identifies" the token(s) satisfying all of the types, the logic infomorphisms are those infomorphisms that transfer token-identification information between local logics, and the channels comprise sets of (logic) infomorphisms encoding mutual constraints that assemble multiple identified tokens. The latter may be naturally viewed as "parts" fitted into a larger identified system (Fields and Glazebrook, 2019b). The logic infomorphisms in principle compose the class of infomorphisms that will drive the inferential mechanism of the information flow in §7. The origin of several of the above concepts may be traced back to the logico-philosophical survey of Barwise and Perry (1983).

### 6.3 Sequents and conditional probabilities

One of the important concepts connected with a local logic is that of a *sequent* as discussed more fully in the Appendix, but we will straightaway formulate this concept in the following:

**Definition 6.3.** A *sequent* $M \Vdash_{\mathcal{A}} N$ *holding of a classifier* $\mathcal{A}$ is a pair of subsets $M, N$ of $\mathrm{Typ}(\mathcal{A})$ such that $\forall\, x \in \mathrm{Tok}(\mathcal{A})$, $x \Vdash_{\mathcal{A}} M \Rightarrow x \Vdash_{\mathcal{A}} N$.

We observe that a sequent encodes a semantic, e.g. causal constraint that in information flow functions effectively as a logical gate. To see how classifiers can hence be interpreted as probabilistic, consider a sequent in $\mathcal{A}$ and its satisfaction condition taken to satisfy:

$$M \Vdash_{\mathcal{A}} N \qquad \forall x (x \Vdash_{\mathcal{A}} M \Rightarrow x \Vdash_{\mathcal{A}} N) \tag{6.5}$$

As described in Allwein (2004); Allwein, Moskowitz and Chang (2004)(see also Fields and Glazebrook (2019a)), to assign a probability to this sequent is to remove the universal quantifier $\forall$, and to assign a probability to $x \Vdash N$, given that $x \Vdash M$, for arbitrary $x$; that is, to assign a probability that $x$ satisfies $N$ given that it satisfies $M$. This is a conditional probability that motivates defining:

$$M \Vdash^{P}_{\mathcal{A}} N := P(M|N) \tag{6.6}$$

This approach is similar to how a conditional probability can be used for interpreting the logical implication "$\Rightarrow$" as discussed in Adams (1998). Accordingly, when the sequent's conditional probability is $p$, we have $M \Vdash^{p}_{\mathcal{A}} N$, noting that we must have $x \Vdash_{\mathcal{A}} M$ in order to apply $M \Vdash_{\mathcal{A}} N$ in a argument. If the probability of the former holding in $\mathcal{A}$ is $P(M)$, then $x \Vdash_{\mathcal{A}} N$ follows from the usual rule $P(M) \cdot M \Vdash^{P}_{\mathcal{A}} N$. Note that $\Vdash$ is a pre-existing concept, which when interpreted probabilistically provides a definition of conditional probability. Instead of being "merely" a map to the real interval $[0, 1]$ as in §2.1, the notion of probability here inherits the semantics associated with the sequent, e.g. a causal or inferential semantics as discussed above. In Bayesian terms, $M$ can be regarded as an unobservable event and $N$ an observable quantity, in which case $P(M)$ is a prior and $P(N)$ is the evidence. Given the likelihood $P(N|M)$, Bayes' theorem specifies (6.6) as the posterior:

$$P(M|N) = \frac{P(N|M)P(M)}{P(N)} \tag{6.7}$$

In Bayesian belief updating ($P(N) = 1$); in this case, (6.7) is a *generative model* $P(M|N) = P(N|M)P(M)$ (recalled in e.g. Kuchling et al. (2019)) that converts a prior to a posterior in each classifier. Information flow via (logic) infomorphisms between classifiers, with sequents '$\vdash$' relaxed at each stage as explained above, provides the necessary semantic consistency for such updating. Infomorphisms thus capture a significant representation of Bayesian inference with the necessary coherence, since the "target" classifier (context) admits the same semantics as that of the "source" within the information flow (cf. McClelland (1998)).

## 6.4 Constructing Cone-Cocone Diagrams

As the first step in constructing a CCCD, we define a finite *information channel* **Chan** as a finite indexed family $\{f_i : \mathcal{A}_i \rightleftarrows \mathbf{C}\}_{i \in \mathcal{I}}$ of infomorphisms having a common core $\mathbf{C}$ as in (6.3). Following §3 the component classifiers $\mathcal{A}_i$ will be taken to characterize a "system of observables", or some sub-components thereof.

In the sense of maximally abstract, while preserving the mathematical structure of interest, the most general channel on a finite set of classifiers corresponds to the category-theoretic notion of a finite *cocone* (the prefix "co-" indicating dual, in this case of a cone), with the core $\mathbf{C}'$ the colimit of all possible upward-going structure-preserving maps from the classifiers $\mathcal{A}_i$ (Adámek, Herrlich and Strecker, 2004; Awodey, 2010). Such a colimit core, provided it exists, can be regarded as "containing" in its structure, as a classifier, all of the information that is common to the component classifiers $\mathcal{A}_i$ (Fields and Glazebrook, 2019a, provides a detailed construction and examples). The Cocone Diagram (CCD), as displayed in (6.8) below, must commute, i.e. the rightward arrows $\mathcal{A}_i \to \mathcal{A}_j$ between the component classifiers must be such that $f_i = f_j g_{ij}$ for all $i, j$, where $g_{ij}$ can be any composition of arrows $\mathcal{A}_i \to \cdots \to \mathcal{A}_j$. This commutativity requirement makes explicit the role of $\mathbf{C}'$ as a "wire" or shared memory for the component classifiers (Barwise and Seligman, 1997; Allwein, Yang and Harrison, 2011). It assures inferential coherence by assuring joint activation of all of the classifiers covered by the cocone core $\mathbf{C}'$, and hence joint, parallel use of all possible inferential paths through the CCD.

$$
\begin{tikzcd}
 & & \mathbf{C}' & & \\
\mathcal{A}_1 \arrow[urr, "f_1"] \arrow[rr, "g_{12}"'] & & \mathcal{A}_2 \arrow[u, "f_2"'] \arrow[rr, "g_{23}"'] & & \dots\ \mathcal{A}_k \arrow[ull, "f_k"']
\end{tikzcd}
\tag{6.8}
$$

A commuting finite cone of infomorphisms is the dual construction, in which all the arrows are reversed. In this case the core of the (dual) channel is the limit of all possible downward-going structure-preserving maps to the classifiers $\mathcal{A}_i$.

We can now define the central idea of a finite, commuting CCCD as comprising both a cone and a cocone on a single finite set of classifiers $\mathcal{A}_i$.

$$
\begin{tikzcd}
 & & \mathbf{C}' & & \\
\mathcal{A}_1 \arrow[urr, "f_1"] \arrow[rr, "g_{12}"', shift right] & & \mathcal{A}_2 \arrow[u, "f_2"'] \arrow[ll, "g_{21}"', shift right] \arrow[rr, "g_{23}"', shift right] & & \dots\ \mathcal{A}_k \arrow[ull, "f_k"'] \arrow[ll, "g_{32}"', shift right] \\
 & & \mathbf{D}' \arrow[ull, "h_1"] \arrow[u, "h_2"] \arrow[urr, "h_k"'] & &
\end{tikzcd}
\tag{6.9}
$$

It is natural to interpret this diagram as depicting a flow of constraint information, represented by the component classifiers, from $D'$ through a set of component classifiers to $C'$; as noted above, this "information flow" is inferential in a natural sense. Commutativity here requires that any path from $D'$ to $C'$, including any number of lateral maps between component classifiers, yields the same result; this assures that all available inferential paths are brought to bear on the "input" encoded by $D'$, and hence assures inferential coherence. The bidirectional maps $\mathcal{A}_i \leftrightarrow \mathcal{A}_j$ between component classifiers are naturally interpreted as encoding mutual, lateral constraints on the behavior of the component classifiers; these are logical constraints imposed by the component classifiers on each other. The maps $f_i$ and $h_j$ can be arbitrarily finitely expanded by inserting intermediate "layers" of additional classifiers, e.g. $f_i : \mathcal{A}_i \to C' \Rightarrow f_i = f_{ib} f_{ia} : \mathcal{A}_i \to \mathcal{B}_i \to C'$ for some intermediate classifier $\mathcal{B}_i$; hence a CCCD can have an arbitrary finite number of layers of classifiers. The structural similarity between CCCDs and recurrent neural network models is obvious. The formal relationship between a cocone diagram and a general feedforward neural network has been developed in detail (Kikuchi et al., 2003); reversing the arrows and superposing yields a recurrent network. As in Appendix §A and §6.3 above, the semantics of the classifiers is inherited. If the component classifiers represent causal "if – then" relations, the inferences implemented are likewise causal.

# 7 Context-compliant Hierarchical Bayesian networks

## 7.1 Channel Theoretic contextuality

Here we will re-formulate the Chu space $\mathcal{A} = (A, r, X)$ and its components $\mathcal{A}_\alpha$ in §3 in terms of classifiers. Thus, with $A$ and $X = B \times R$ as defined in i)-iii) of §3, consider the classifier:

$$\mathcal{A} = \langle A, X, \Vdash_{\mathcal{A}} \rangle \tag{7.1}$$

where, for now, $\Vdash_{\mathcal{A}}$ is an abstract classification which is to be realized below. Taking component classifiers $\mathcal{A}_\alpha$ as in §3 as comprising *observables in context*, and assuming, without loss of generality, that these are finitely structured, leads to an information flow:

$$\leftrightarrows \mathcal{A}_{\alpha+1} \leftrightarrows \mathcal{A}_\alpha \leftrightarrows \mathcal{A}_{\alpha-1} \leftrightarrows \cdots \tag{7.2}$$

The semantic content can be extended by postulating local logics $\mathcal{L}_\alpha = \mathcal{L}(\mathcal{A}_\alpha)$ generated by the corresponding classifiers $\mathcal{A}_\alpha$, as pointed out in §6.2, and specified formally in Definition A.4 *et seq.* This particular formulation is assumed, in principle, to be in relationship to a (regular) theory associated to the individual $\mathcal{A}_\alpha$ (Definition A.2). Accordingly, a flow of logic infomorphisms:

$$\cdots \leftrightarrows \mathcal{L}_{\alpha+1} \leftrightarrows \mathcal{L}_\alpha \leftrightarrows \mathcal{L}_{\alpha-1} \leftrightarrows \cdots \tag{7.3}$$

can be postulated, and granted this is the case, we can construct a CCD as in (6.8) that depicts this flow. Further, on weakening the sequents of the local logic, as described in §6.3, the individual (abstract) classifications $\Vdash_{\mathcal{A}_\alpha}$ are now explicitly realized as conditional probabilities $p_\alpha(\cdot|\cdot)$. Hence with the inclusion of conditionals, the information flow in the CCD now represents an inferential process. As described in §6.2, the existence constraint on the CCD is, effectively, commutativity.

We are now in a position to give the maps $f_\alpha$ and the associated logic $\mathcal{L}$ an interpretation, and to make this all the more specific. Firstly, commutativity requires that branching "upward" to $\mathcal{L}$ along any one of the $f_\alpha$ is equivalent to following the whole inferential sequence (7.3) to its end, and then following the last of the $f_\alpha$ to $\mathcal{L}$. The $f_\alpha$ are, therefore, shortcuts to reaching $\mathcal{L}$: "insights" that allow the rest of the inferences in (7.3) to be bypassed. The logic $\mathcal{L}$ is the "answer" to the problem (7.3) addresses: formally, it is the logic that solves the problem in one step. The probability that the answer is "right" is the product of the probabilities along (7.3). Secondly, commutativity requires that this overall probability is conserved; hence the probability associated with each "insight" $f_\alpha$ must be the combined probability of the inferential steps it replaces. As in the case of constraint flow in §6.2 above, commutativity or equivalently, the existence of a CCD, enforces inferential coherence. The same clearly applies when the dual objects and maps are added to construct a CCCD.

### 7.2 Commutativity requires co-deployable observables

The question raised and deferred at the end of §3 can now be addressed. A context as described in CbD, and in related theories, that can be represented by elements of the set $R$ in §3, is a set of observations, or in quantum-theoretic terms, outcomes obtained from a set of "deployed observables." In practice, such sets are always finite, and their elements – the "observations" or "outcomes" themselves – are always encodable as finite bit strings. As discussed in Fields and Marcianò (2019b), some of these observables encode reference frames that specify both what is being observed and how it is being observed. A context change is a change in the observables deployed. The observables deployed within any given context must be co-deployable, i.e. the operators representing them must mutually commute. Observables in different contexts do not have to be co-deployable, i.e. their operators need not commute. This lack of commutativity is often regarded as intrinsically "quantum" but in fact arises even in classical theories when the necessarily-finite resolution of practical observations is taken into account (Jennings and Leifer, 2016).

Observational outcomes obtained with sets of co-deployable observables admit well-defined probability distributions; this is guaranteed in quantum theory by the Born Rule and is simply assumed in classical theories. Observational outcomes obtained in non-overlapping contexts are *prima facie* statistically unrelated, though as discussed in §5.2 they may be discovered *post facto* to be statistically related. What happens, however, in partially-overlapping contexts? This is the situation addressed by the contextuality theorems of quantum theory (Bell (1966); Kochen and Specker (1967); see also e.g. Abramsky and Brandenburger (2011)) and is one motivation for CbD and other theories of contextuality.

### 7.3 A Channel Theoretic interpretation of intrinsic contextuality

The formulism of CCDs allows us to formulate this question concerning overlapping contexts with a diagram. Let $\mathcal{A}_1, \mathcal{A}_2, \dots\ \mathcal{B}$ and $\mathcal{B}, \mathcal{C}_1, \dots\ \mathcal{C}_k$ be finite sets of co-deployable observables, as in §7.1, and let $\mathbf{C_1}$ and $\mathbf{C_2}$ be the respective cocone cores of the contexts they define. The combined set $\mathcal{A}_1, \mathcal{A}_2, \dots\ \mathcal{B}, \mathcal{C}_1, \dots\ \mathcal{C}_k$ of observables is co-deployable, in the sense of having a well-defined joint probability distribution, if and only if a core $\mathbf{C}$ and maps $\phi$ and $\psi$ exist such that the following diagram commutes, i.e. is a CCD:

$$\begin{tikzcd}
 & & \mathbf{C} & & \\
 & \mathbf{C_1} \arrow[ur, "\phi"] & & \mathbf{C_2} \arrow[ul, "\psi"'] & \\
\mathcal{A}_1 \arrow[ur, "f_1"] \arrow[r] & \mathcal{A}_2 \arrow[u, "f_2"'] \arrow[r] & \dots\ \mathcal{B} \arrow[ul, "f_B"'] \arrow[ur, "g_B"] \arrow[r] & \mathcal{C}_1 \arrow[u, "g_2"'] \arrow[r] & \dots\ \mathcal{C}_k \arrow[ul, "g_k"']
\end{tikzcd} \tag{7.4}$$

Failure of commutativity at the diagram level is failure of co-deployability, i.e. failure of commutativity at the operator level. We summarize matters in the following:

**Definition 7.1.** With respect to a diagram of the form (7.4), we say that a set of observables deployed within any given context are *co-deployable* if the operators representing them commute. Otherwise, the observables are *non-co-deployable*, and the measurement system comprising them is said to be *intrinsically contextual.*

The requirement that a CCD exists for any set of co-deployable operators has a straightforward interpretation in terms of probability distributions, and leads to the following:

**Theorem 7.1.** *With respect to a diagram of the form* (7.4)*, failure of commutativity is equivalent to the non-existence of a consistent probability distribution across the combined set of observables* $\{\mathcal{A}_i\}$.

*Proof.* Without loss of generality, we assume that the CCDs below $\mathbf{C_1}$ and $\mathbf{C_2}$ encode well-defined probability distributions as described above. Failure of commutativity in (7.4) then implies either that $\phi$ does not have the combined probability of the horizontal maps from $\mathcal{B}$ to $\mathcal{C}_k$, or that $\psi$ does not have unit probability. Either way, no well-defined probability distribution across the combined set of observables exists. Conversely, if no probability distribution on the combined set of observables exists, then the probability of at least one of $\phi$ and $\psi$ is undefined and hence the CCD does not exist. □

**Example 7.1. Calibration and measurement**

A simple example of induced contextuality is provided by the commonplace practice of calibrating an instrument (Fields, 2018). Let $\mathcal{A}_1 \dots \mathcal{A}_n$ be a set of classifiers that identify a particular instrument, i.e. distinguish it from other items in a laboratory. Let $M_p$ be an operator, considered as a classifier, that measures the value of a "pointer" $p$ and let $C_p$ be an operator that calibrates $p$ with respect to some standard. Clearly $M_p$ and $C_P$ in general do not commute, i.e. $[M_p, C_p] \neq 0$. If we consider a diagram:

$$\begin{tikzcd}
 & & & \mathbf{C} & & & \\
 & \mathbf{C_1} \arrow[urr, "\phi"] & & & & \mathbf{C_2} \arrow[ull, "\psi"'] & \\
 \mathcal{A}_1 \arrow[r] \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[u, "f_2"'] & \dots \arrow[r] & M_p \arrow[ull, "g_1"'] & \mathcal{A}_1 \arrow[r] \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[r] \arrow[u, "f_2"'] & \dots\ C_p \arrow[ul, "g_2"']
\end{tikzcd} \tag{7.5}$$

the component cocone core $\mathbf{C_1}$ representing the combined operations of identifying the instrument and measuring the pointer will fail to commute with $\mathbf{C_2}$ representing the combined operations of identifying the instrument and calibrating the pointer. Thus the overall diagram will fail to commute. In this case, by Theorem 7.1, the probability distribution defined by the maps $\phi$ and $\psi$ is not self consistent; indeed, it violates the Leggett-Garg inequality when measurement and calibration are viewed as sequential (Fields, 2018).

**Example 7.2. System identification and "settings"**

We can also consider the complement of the above example, in which contextuality is induced by a "swap" of one system for another, functionally-equivalent system. Such swaps may be undetected or undetectable, e.g. if the two systems are indistinguishable with the available system-identification operators, except *post facto* as discussed above (Fields, 2018). Here let $\mathcal{A}_1 \dots \mathcal{A}_n$ be a set of classifiers that identify a particular instrument, and $\mathcal{B}_1 \dots \mathcal{B}_n$ be classifiers that identify a numerically (or "ontologically") distinct instrument. Let $M_p$ measure the value of the "pointer" $p$ of the first instrument and let $M_q$ measure the pointer $q$ of the second instrument. Clearly $M_p$ and $M_q$ commute, i.e. $[M_p, M_q] = 0$, as they are acting on distinct physical degrees of freedom. The relevant diagram is:

$$\begin{tikzcd}
 & & & \mathbf{C} & & & & \\
 & \mathbf{C_1} \arrow[urr, "\phi"] & & & & \mathbf{C_2} \arrow[ull, "\psi"'] & & \\
\mathcal{A}_1 \arrow[r] \arrow[ur, "f_1"] & \mathcal{A}_2 \arrow[u, "f_2"'] & \dots \arrow[r] & M_p \arrow[ull, "g_1"'] & \mathcal{B}_1 \arrow[r] \arrow[ur, "h_1"] & \mathcal{B}_2 \arrow[u, "h_2"'] \arrow[r] & \dots & M_q \arrow[ull, "g_2"']
\end{tikzcd} \tag{7.6}$$

If the "observer" in this case can distinguish the $\mathcal{A}_1 \dots \mathcal{A}_n$ from the $\mathcal{B}_1 \dots \mathcal{B}_n$, and hence distinguish the first instrument from the second, this diagram depicts a straightforward case of direct influence: using a different instrument, even one of "identical" type and characteristics, can be expected to yield a different result, or more generally, to yield results with a distinct probability distribution. If, however, we introduce a no-signalling condition that prevents the observer from distinguishing the $\mathcal{A}_1 \dots \mathcal{A}_n$ from the $\mathcal{B}_1 \dots \mathcal{B}_n$, the context switch goes unnoticed (is indeed not noticeable, by assumption), and the situation is contextual. No signalling can be implemented here by having Alice identify the instruments and Bob record the measurement outcomes while preventing communication between the two. A similar situation results if Alice adjusts the "settings" on a single instrument (e.g. the paired Stern-Gerlach devices in a Bell/EPR experiment) while Bob records the results, provided the two cannot communicate.

**Example 7.3. The "Snow Queen" experiment**

The example above illustrates the importance of "signalling" between observers or instances of observation in distinguishing between direct influences and intrinsic contextuality (cf. §4.2). We can also consider signalling between two parts of a single observer's cognitive system. As Cervantes and Dhzafarov (2018) point out in the analysis of their "Snow Queen" experiment, intrinsic contextuality results when what appear to be single "concepts" (e.g. 'kind' or 'beautiful') have different meanings in different contexts, and these differences in meaning extend beyond those induced by the given descriptions of the contexts (i.e. beyond what is actually "observed" in each context). Differences in meaning are, effectively, restrictions on communication: conflicting alternative meanings for a single word or concept reflect the absence or dysfunction of an error correction system capable of accessing and resolving the conflict. Differences in meaning between contexts lead to differences between contexts in what can be inferred or put into practice (e.g. Roederer, 2005). When meanings are distinct enough between contexts to be mutually incoherent, actions appropriate to one meaning (in the "Snow Queen" experiment, a meaning of "beautiful" that incorporates emotional characteristics such as kindness) are inappropriate (given constant goals) to the other (e.g. a meaning of "beautiful" based entirely on physical features). Hence one does not want to fall in love with the Snow Queen.

We now state a general channel–theoretic version of Example 4.1:

**Corollary 7.1.** *Let* (7.4) *represent a measurement system in which, for some $i, j$, subsets $\{a_i\} \subset A$ of events, and conditions $\{b_j\} \subset B$ are measured in every context in $R$. Assume that diagram* (7.4) *fails to commute. Then there exists at least one pair $c, c'$ of contexts in $R$ with $c \neq c'$, and at least one event $a \in \{a_i\}$, and condition $b \in \{b_j\}$, such that $p(a|\{b, c\}) \neq p(a, |\{b, c'\})$.*

*Proof.* Consider the following classifiers in a non-commuting sector of (7.4). Taking a subset of contexts $\{c_k\} \subset R$, for some $k$, and a corresponding subset $\{x_{jk}\} = \{b_j, c_k\} \subset X = B \times R$, leads to a component classifier $\mathcal{A}_{ijk} = \langle \{a_i\}, \{x_{jk}\}, \Vdash_{ijk} \rangle$. Let $\mathcal{A}$ be the subcomponent of $\mathcal{A}_{ijk}$ consisting of $a \in \{a_i\},\ b \in \{b_j\},\ c \in \{c_k\}$, and $\mathcal{A}'$ that consisting of $a, b, c' \neq c$. In such a non-commuting sector

$$\begin{tikzcd} & \mathcal{A}_{ijk} & \\ \mathcal{A} \arrow[ur, dotted] \arrow[rr, dotted, dash] & & \mathcal{A}' \arrow[ul, dotted] \end{tikzcd} \tag{7.7}$$

Theorem 7.1 implies that there is no definable consistently shared probability distribution in this sector. In particular, the horizontal dotted line between $\mathcal{A}$ and $\mathcal{A}'$ in (7.7), indicating no such consistent probability distribution exists between these two classifiers, implies the result. □

"Intrinsic contextuality" is, in other words, a failure of commutativity; a set of observables $\mathcal{A}_1, \mathcal{A}_2, \ldots\ \mathcal{B}, \mathcal{C}_1, \ldots\ \mathcal{C}_k$ exhibits intrinsic contextuality if a cocone does not exist above them. The component cocones, in this case, those under $\mathbf{C_1}$ and $\mathbf{C_2}$, are in this case mutually incoherent, i.e. the corresponding measurement systems fail to be co-deployable. We can think, in this representation, about the process of incrementally adding to or replacing the observables being deployed in a given situation. Starting with a coherent set over which a cocone (and hence a CCD) is defined, adding or replacing observables merely changes the definition of the cocone as long as the new observables are co-deployable with the original ones. The altered set "poses a coherent problem" for which the new cocone core is the one-step solution. Adding or replacing co-deployable observables introduces "direct influences" on the probabilities of single-observable outcomes by changing the problem that the observations are an attempt to solve. Such changes can lead to interpretative difficulties if the change in problem is not recognized as such, but do not lead to incoherence. As long as the observables are co-deployable, it is possible to treat all of the variables in the context as one composite random variable ($R_q^C$), as discussed in (Dzhafarov, Kujala and Cervantes, 2016a; Dzhafarov, Cervantes and Kujala, 2017b) and in §3.1. Introducing non-co-deployable observables, however, introduces incoherence. This is the situation, for example, in an artificial neural network (ANN) with an inconsistent weight matrix. This kind of inconsistent scenario would correspond to an ANN that fell short of a unique answer and instead engaged in parasitic oscillations, or to a categorization network that activated two or more mutually-incoherent categories depending on what input nodes were active (cf. the "fractured co-regulators" of Ehresmann and Gomez-Ramirez (2015)). The above approach (including Theorem 7.1) may thus be compared with the non-consistently connected interpretation of contextuality discussed in §4.1, and with the non-existence of global sections of a sheaf of certain distributions over a "measurement cover" as characterizing contextuality and non-locality in Abramsky and Brandenburger (2011). In this respect, the failure of commutativity hypothesized in Theorem 7.1 leading to non-existence of an overall consistent probability distribution, appears to be compatible with, and indeed to provide a conceptual underpinning for these more probability-theoretic approaches.

### 7.4 A signed CCCD

Let us now return to the space of signed probability measures $\mathcal{M}(Z)$ in §5.3, and assume that $A, B, X = B \times R$ are subsets of $\mathcal{M}(Z)$. In doing this we obtain a system of signed component classifiers $\mathcal{A}^{\pm}_{\alpha}$, where the condition that $A, B, X$ are subsets of $\mathcal{M}(Z)$, leads to some scope in how these arguments can be signed. For instance, $A = A^{\pm}$ could be signed events, such as those in §2.6 equipped with a signed probability measures, and $\Vdash_{\mathcal{A}_\alpha}$ corresponds to a continuous function $p : A \times X \longrightarrow [-1, 1]$. Much of the above considerations apply, and adapting diagram (6.8), we arrive at a signed CCD

$$\begin{tikzcd} & \mathbf{C}^{\pm} & \\ \mathcal{A}_1^{\pm} \arrow[ur, "f_1^{\pm}"] \arrow[r, "g_{12}^{\pm}"'] & \mathcal{A}_2^{\pm} \arrow[u, "f_2^{\pm}"'] \arrow[r, "g_{23}^{\pm}"'] & \cdots\ \mathcal{A}_k^{\pm} \arrow[ul, "f_k^{\pm}"'] \end{tikzcd} \tag{7.8}$$

Moreover, (7.8) can be fed into (7.4) to create a signed version of the latter that now houses negative probabilities, and combining signed (7.4) with its dual yields a signed version of (6.9), i.e. a signed CCCD. Again relaxing the sequents as in §6.3, we obtain an inferential system of conditional probabilities, when defined, through a signed information flow. The argument at the end of §7.1 still applies: the signed probability distribution must be well-defined if the signed CCD or the signed CCCD is to exist. In particular, Theorem 7.1 and Corollary 7.1 have their signed counterparts. For instance in quantum correlation studies, involving observables deployed within a context or with respect to a reference frame, the commutativity or non-commutativity of signed (7.4) with its negative probability measures would amount to violating the tenets of, e.g. local deterministic hidden variable models, and hence the Bell inequalities which require classical probabilities (Fine, 1982). Commutativity of signed (7.4) may be compared with the result of Abramsky and Brandenburger (2011, 2014) that an empirical model is no-signalling, if and only if it admits a global sheaf section of distributions over a measurement cover. In this regard, probability models have local hidden variable realizations with respect to negative probabilities, if and only if they satisfy no-signalling (Abramsky and Brandenburger, 2011, Th. 5.9).

## 8 Context-compliant Hierarchical Bayesian inference and contexts in practice

### 8.1 The Bayesian picture

Cocones are naturally interpreted as encoding abstraction (Goguen, 1991), or in psychological terms, categorization by abstracted types. Fields and Glazebrook (2019b) shows how to construct visual categorization hierarchies, beginning with "object files" representing located, bounded, featured entities that are visibly distinct from their accompanying backgrounds (Kahneman, Triesman and Gibbs (1992); see also Flombaum, Scholl and Santos (2008); Fields (2012)). In the dual, top-down direction, individual objects encode consistency conditions over abstract categories; a black cat, for example, demonstrates the consistency of the categories 'animal', 'four-legged thing', and 'black thing'. Indexing a set of object files $\{Y_i\}$ by time allows construction of time-indexed classifiers $\mathcal{O}_{i,t} = \langle \{O_{i,t}\}, \{Y_i\}, \Vdash_P \rangle$, where $\{O_{i,t}\}$ is a time-indexed sequence of identified objects and $\Vdash_P$ denotes an empirically determined relation consistent with object permanence (Fields and

Glazebrook, 2019b). This construction supports the representation of "object tokens" representing re-identifiable objects with histories, and hence supports the representation of feature change. Dually, time-extended objects encode consistency conditions on sequences of feature changes.

Fields and Glazebrook (2019b) also introduce an interpretation of cocones as representations of mereological (i.e. part-whole) inclusion. Mereological hierarchies are logically orthogonal to abstraction hierarchies and have, curiously, received far less attention in the cognitive neuroscience literature. As in abstraction hierarchies, lower-level entities serve the dual role of encoding consistency constraints across the higher-level entities that contain them. Employing CCCDs as representations of mereological classification enables not just the assembly of parts (e.g. the face, legs, tail etc. of a cat) into wholes, but also the assembly of objects into scenes and the assembly of object tokens with histories into sequences of episodic memories.

As discussed in Fields and Glazebrook (2019b), Bayesian brain models represent both type categorization and mereological assembly as Bayesian inferences regulated by prior probability distributions. It is standard to view prior probabilities as "downward" constraints that assemble sets of expected lower-level features; this interpretation of prior probabilities is implemented by predictive-coding models (e.g. Knill and Pouget, 2004; Bubic, von Cramon and Schubotz, 2010; Clark, 2013; Friston and Kiebel, 2009). An alternative interpretation is suggested by (7.3) and the probabilistic interpretation of CCDs: prior probabilities are "shortcuts" that allow "jumping to the conclusion" of an abstract categorization or mereological assembly, with the implications for actionability that such conclusions provide. This interpretation points toward the obvious evolutionary advantage of object- and categorical type-memories: they increase the efficiency of problem solving (i.e. appropriate action selection) in a dynamic, resource-limited environment.

Within this Bayesian picture, mutually-incoherent cocones lead to incoherent object identification, incoherent categorization and/or mereological assembly, and incoherent inferences about appropriate action. Intrinsic contextuality, in other words, provokes inferential incoherence, as Examples 7.1, 7.2 and 7.3 illustrate. How can a hierarchical Bayesian classifier cope with intrinsic contextuality? Here the definition of a context as *what is observed* shows both its strength and its weaknesses. The effects of a context can be better understood, and hence predicted, by expanding it: embedding it in a larger context by observing more. In the limit, direct influences become predictable, so no longer surprising. A context can, however, only be expanded by deploying additional co-deployable observables. If not all available observables are co-deployable, mutually-incoherent contexts and hence intrinsic contextuality will remain. The challenge for a Bayesian classifier is, then, to be able to recognize when observables are not co-deployable, or equivalently, when concepts like "beautiful" in the "Snow Queen" experiment (Example 7.3) switch between mutually-incoherent meanings. Quantum theory formalizes the needed knowledge in Bohr's notion of complementarity: observables are complementary whenever their operators do not commute (Bohr, 1928). Moving beyond a formalized theory, however, is challenging. Observers cannot, in general, determine by observation what observables they are deploying (Fields and Marcianò, 2019b). "Learning the hard way" by working backwards from outcomes that reveal the effects of incoherent inferences is the fallback option (again cf. the discussion in §4.2); indeed this is how complementarity was discovered by physics. Working backwards requires memory, time, and other resources that can be in short supply in a dynamic environment.

### 8.2 Contexts in practice I: The idea of active inference

As Friston (2010, 2013) has emphasized, observation is not a passive process (c.f. enactive cognition, e.g. Froese and Ziemke, 2009). Observers constantly probe their environments through action, influenced by hidden states, and time-sensitive policies, when seeking to decrease variational free energy by modifying environmental states in meeting with their expectations, as to their advantage. In this respect, active inference may be viewed as a paradigm for contextual learning. Grasping and moving a coffee cup for an enjoyable sip, is a very simple example.

What actions, however, are appropriate? What actions will reveal aspects of the environment important for defining and reasoning within a context? What actions will reveal that a context has changed? Here a context *having changed* entails, by definition, that the observables deployed by the observer have changed, generally without the observer's knowledge or intent. Hence the question is the one of "working backwards" raised above: what actions will reveal, by their effects, what observables are in fact being deployed?

Questions like these make little sense for ideally rational observers embedded in fully-transparent environments, e.g. the perfectly efficient markets of classical decision theory (e.g. Parmigiani and Inoue, 2009). They arise immediately, however, when the observer is separated from the world by a Markov Blanket (Friston, 2013; Clark, 2017) or, equivalently, an interface (Hoffman, Singh and Prakash, 2015) that renders observations conditionally independent from the world, including the effects of actions. For such an observer, observational outcomes are effectively encoded by the blanket. These blanket-encoded outcomes are the only information about the environment that is available. A context change, in this case, is simply the appearance of new observational outcomes. Nothing *principled* distinguishes a context change from the evolution of a fixed context. Consequences of actions on the environment, in particular, are only deducible indirectly from observations, i.e. consequences of the environment's action back on the observer, consequences that may appear only well after the original action and not be associated with it. Here again, the question of the observer's memory and inferential resources becomes important. To see this in terms of the CCCD architecture, in regards to diagrams (7.4)(7.6) say, we note that the bottom level of classifiers corresponding to those exposed to the world, effectively functions as a Markov blanket. The bidirectional, physical "measurement" operators corresponding to these classifiers in Fields and Glazebrook (2020a), which implement both input from and output to the observed "world" implement active inference with respect to the corresponding Bayesian (i.e. CCD) configuration.

### 8.3 Contexts in practice II: The Frame Problem

The situation faced by such an observer as above is well known in AI, where it is formulated by the *Frame Problem*: the problem of circumscribing what does *not* change when an action is performed (McCarthy and Hayes, 1969; Dietrich and Fields, 1996; Shanahan, 2016). Viewed broadly, it is the problem of circumscribing what is relevant in a situation. As solving the Frame Problem requires, in effect, checking all facts in a situation to look for changes, it rapidly becomes unsolvable in practice, except approximately through the use of greedy heuristics, as the domain of action becomes large (indeed the general case is formally undecidable (Dietrich and Fields, 2020)). The problem of individual object re-identification over time can be recast as an instance of the Frame Problem (Fields, 2013b); this problem is only "solved" heuristically. [††]

[††] The Frame Problem has been popularized as a form of Hamlet's problem: when to stop thinking? It describes the predicament that Fodor's hapless robot falls victim to (Dennett, 1984; Shanahan and Baars, 2005)(cf. Fodor

Heuristic solutions to the Frame Problem assume a coherent context; they assume that what is observed could, in principle, be expanded to include all that there is to be observed. The infeasibility of such an expansion with limited resources is the reason heuristics are required.[‡‡] In the presence of intrinsic contextuality, however, this coherence assumption is violated. If there is intrinsic contextuality, i.e. if some observables are not co-deployable, the current context cannot, even in principle, be expanded to include everything that is observable. To take a very simple example, measurement contexts for position observables cannot be extended to include momentum observables, as these do not commute. Hence values of momentum can only be updated heuristically following measurements of position. Intrinsic contextuality, therefore, renders the Frame Problem unsolvable even in "small" domains.

To reconstruct the Frame Problem in our formulism, we introduce the idea of an *informationally unencapsulated process* (Shanahan and Baars, 2005) employed in cognitive science and AI. An informationally unencapsulated process effectively assimilates all information from any source that is *relevant in fact* for solving a problem. Such a process is not limited by *a priori* circumscriptions of what is relevant, and is taken, at least in principle, to be unlimited by resource availability. As an informationally unencapsulated process "knows" all relevant consequences of an action, it can solve the frame problem. The question of interest in then whether such processes exist.

Given some distributed system of information flow as in Example 6.3, one can ask whether further infomorphisms $\phi$ and $\psi$ can be found to complete a cocone as in (7.4). Is it possible, in other words, to construct a commutative diagram:

$$\begin{tikzcd}
 & & ? & & \\
 & \mathbf{B}_1 \arrow[ur, dotted, "?"] & & \mathbf{B}_2 \arrow[ul, dotted, "?"'] & \\
\mathcal{A}_1 \arrow[ur, "f_1"] & & \mathcal{A}_2 \arrow[ul, "f_2"'] \arrow[ur, "f_3"] & & \mathcal{A}_3 \arrow[ul, "f_4"']
\end{tikzcd} \tag{8.1}$$

As the colimit classifier $\mathbf{C}$ that completes such a diagram captures all of the information from the underlying classifiers $\mathcal{A}_i$, the completed diagram represents an informationally unencapsulated process if, but only if, the $\mathcal{A}_i$ gather all of the information that is *in fact* relevant to whatever problem the distributed system is meant to solve. The conditions that must be met for this to be the case are given by Theorem 7.1: the colimit exists only if the $\mathcal{A}_i$ are co-deployable. Hence we have:

**Corollary 8.1.** *A distributed information-flow system can be informationally unencapsulated only in the absence of intrinsic contextuality.*

From a practical standpoint, proving the absence of intrinsic contextuality in a domain (in AI terms, a task environment) requires discovering all of the information that is relevant in fact to

(1983)). The difficulty, of course, is that for any informationally unencapsulated inferential process, no *a priori* limit exists to information pertinent to that process.

[‡‡] At least as far as "heuristic solutions" for humans go, such solutions emerge from the methodology of the massively parallel, competitively based, distributed system implemented by the GNW architecture (Dehaene and Naccache, 2001; Shanahan and Baars, 2005). Indeed, the GNW can be seen as a mechanism for determining relevance on the fly.

solving problems in that domain. Hence the Frame Problem is solvable only if it has already been solved.

# 9 Conclusion

"Context" is often used to designate what is neglected when an observation is made or an action undertaken; "environment" is often used in a similar way. The formalism of a theory such as CbD challenges us to acknowledge that a context is always there, whether attended to or not. By defining "context" as whatever is observed, this emphasizes that every aspect of a situation is "context" for every other aspect. As Dzhafarov, Kujala and Cervantes (2016b) show, most "contextual" effects in psychology, and by extension, in biology and other sciences of complex systems, are direct influences of context that become explicable as the context is expanded by including more observations. However, a core of intrinsic contextuality remains, both in physics where it was first characterized (Bell, 1966; Kochen and Specker, 1967), and in psychology (Cervantes and Dhzafarov, 2018; Basieva et al., 2019).

Here we have formulated an approach to intrinsic contextuality in general category-theoretic terms: a set of observations exhibits intrinsic contextuality if no cocone can be constructed over the observables that produced them. Mutually incoherent cocones over subsets of observables indicate the inability to impose a mathematically consistent joint probability distribution over the combined subsets. This reformulation is scale-free and applies to both quantum and classical systems. In a hierarchical Bayesian setting, intrinsic contextuality leads to mutually-incoherent classifications and hence mutually-incoherent inferential paths. In the presence of intrinsic contextuality, indicated by non-co-deployable operators as in Definition 7.1, the Frame Problem becomes unsolvable in principle even in "small" domains.

The current work raises a number of questions and open problems. The first, already posed forcefully by Basieva et al. (2019), is the further exploration of intrinsic contextuality in humans by standard experimental means. A second, however, is the question of how switching between mutually-incoherent contexts is implemented at the neurocognitive level. What attentional processes detect the inconsistencies induced by intrinsic contextuality, and how are prior probabilities updated in response? The existence of complementary observables has perplexed physicists for decades (Feynman characterized it as "the only mystery" of quantum theory; see Feynman, Leighton and Sands (1965), Vol. III, §1.1). How are such paradoxes resolved at the implementation level, if they are, or worked around if they are not? What happens when work-around mechanisms fail? What is required to detect failure *post hoc*? An ability to probe this experimentally would begin to answer the broader question of how "prior probabilities" are represented in memory, and how different sets of priors can be deployed in different contexts.

A third and broader question is whether intrinsic contextuality can be detected and characterized at multiple scales in complex systems generally. By formulating a scale-free model of contextuality in category-theoretic terms, and in particular employing the concepts of classifiers and information channels, we hope to have opened a new pathway towards addressing these questions in multiple systems.

# A Appendix: Theories and Local Logics in Channel Theory

In this Appendix we summarize the essential concepts and results pertaining to theories, local logics, and the logic infomorphisms of Channel Theory, referring to Barwise and Seligman (1997) for details (see also Fields and Glazebrook (2019a); Seligman (2009)). We commence with some set $\Xi$, which could be viewed as set of "types", with a (binary) consequence relation $\vdash$ between subsets of $\Xi$. A *sequent* is a pair $I = \langle \Gamma, \Delta \rangle$ of subsets of $\Xi$ (it is sometimes useful to view $\Gamma$ and $\Delta$ 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 $\Gamma$, then it supports some type in $\Delta$.

**Definition A.1.** A *theory* is a pair $T = \langle \Xi, \vdash_T \rangle$, where $\vdash_T$ is a consequence relation on $\Xi$. A *constraint* of the theory $T$ is a sequent $\langle \Gamma, \Delta \rangle$ of $\Xi$ for which $\Gamma \vdash_T \Delta$. A a sequent $\langle \Gamma, \Delta \rangle$ is said to be *$T$-consistent* if $\Gamma \nvdash_T \Delta$.

**Example A.1.** A simple example illustrates this principle. Suppose $\Xi$ is the set of polynomials in two variables $x, y$, and let $\vdash$ be the consequence relation consisting of sequents $\langle \Gamma, \Delta \rangle$, such that each pair $(u, v) \in \mathbb{R}^2$ satisfying all equations in $\Gamma$, satisfies some equation in $\Delta$. For instance, a constraint of the theory might be $x^2 + y^2 = 25,\ 3x = 4y \vdash x = \pm 4$ (Barwise and Seligman, 1997, Ex 9.2, p.118).

In practice, conceiving some aspects of a situation, either causally requiring or merely causally allowing other aspects of a situation, makes the above definition clear.

Every classifier has a theory associated with it in the following way (see Barwise and Seligman (1997, Prop 9.5)).

**Definition A.2.** A *theory* $\mathsf{Th}(\mathcal{A}) = (\Sigma_{\mathcal{A}}, \vdash_{\mathcal{A}})$ *generated by a classification* $\mathcal{A}$, is one that satisfies for all types $\alpha$ and all sets $\Gamma, \Gamma', \Delta, \Delta', \Sigma', \Sigma_0, \Sigma_1$ of types:

(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 $\Sigma$, then $\Gamma \vdash \Delta$.

More generally, any theory $T = \langle \Sigma, \vdash_T \rangle$ is said to be *regular* if it satisfies the above three conditions.

Similar to how classifiers admit infomorphisms, theories have their own notion of morphism, namely, a "theory interpretation" (Barwise and Seligman, 1997, §9.4). Given a theory $T$, let Typ(T) be its set of types, and $\vdash_T$ its consequence relation.

**Definition A.3.** A *(regular) theory interpretation* $f : T_1 \longrightarrow T_2$ is a function from $\mathrm{Typ(T_1)}$ to $\mathrm{Typ(T_2)}$, such that for each $\Gamma, \Delta \subseteq \mathrm{Typ(T_1)}$, if $\Gamma \vdash_{T_1} \Delta$, then $f[\Gamma] \vdash_{T_2} f[\Delta]$ (here "$f[\ ]$" denotes the image of the theory under the function).

**Definition A.4.** *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 classifier $\mathcal{L} = \langle \mathrm{Tok}(\mathcal{L}), \mathrm{Typ}(\mathcal{L}), \Vdash_{\mathcal{L}} \rangle$,

(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 are tokens that satisfy all of the constraints of the theory $\mathrm{Th}(\mathcal{L})$ in (2).

For any classifier $\mathcal{A}$ there is the natural local logic $\mathsf{Lg}(\mathcal{A})$ generated by $\mathcal{A}$: it has classifier $\mathcal{A}$, a regular theory $\mathsf{Th}(\mathcal{A}) = (\mathrm{Typ}(\mathcal{A}), \vdash_{\mathcal{A}})$, and all of its tokens are normal. 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). Intuitively, a local logic is "local" to the classifier that generates it. Infomorphisms allow mapping the local logic of one classifier to that of another; hence we can think of channels as supporting the flow of locally-defined logical relations between classifiers. In Fields and Glazebrook (2019a) it was noted that any classifier can be interpreted as defining a coarse-graining, and hence a "scale" at which information is being organized and represented. Thus each local logic can be thought of as a logic at some level of description.

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 (c.f. Friston (2010)). These observations lead to asking to what extent an infomorphism between classifiers (denoted $\mathsf{Cl}$) will respect their respective local logics. This is specified by the following (Barwise and Seligman (1997, 12.3)):

**Definition A.5.** 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 classifiers.

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

(3) $f^{\vee}[\mathsf{N}_{\mathcal{L}_2}] \longrightarrow \mathsf{N}_{\mathcal{L}_1}$ (Comment: the notation is that of Barwise and Seligman (1997, Definition 12.16). The "[ ]" notation is explained in Definition A.3 above. By covariance of $f^{\vee}$, (3) is equivalent to the pushforward $f^{\vee}(\mathsf{N}_{\mathcal{L}_2}) \subseteq \mathsf{N}_{\mathcal{L}_1}$.)

Consequences of this definition are outlined in Barwise and Seligman (1997, §12.3) (for an application of the above concepts to graded consequences and belief, see Dutta, Skowron and Chakraborty (2019)), but we again emphasize that the flow of information through a network of logic infomorphisms is naturally interpretable as "inference" in the usual sense. As the classification $\Vdash$ can be considered time- and context-dependent, these inferential processes can be regarded as having similar dependencies. Note that local logics as defined above, are implemented by classifiers, and not by maps between e.g. neurons (perceptrons), or co-activated functional assemblies of such. This difference in definition does not rule out the possibility that local logics are themselves implemented by co-activated functional assemblies of neurons, that which Ehresmann and Vanbremeersch (2007) call "cat-neurons.". Hence they can be considered as logic infomorphisms (see Fields and Glazebrook (2019a,b) for examples).

## Conflict of interest

The authors report no conflicts of interest involved in this work.

## Acknowledgement

The authors wish to thank several anonymous referees for their constructive criticism, and helpful comments which assisted in improving the overall presentation of this paper. CF thanks the Federico and Elvia Faggin Foundation for financial support.

## References

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

Abramsky, S. & Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics* 13, 113036

Abramsky, S. & Brandenburger, A. (2014). An operational interpretation of negative probabilities and no-signaling models. In (van Bruegel, F. et al. eds) Panangaden Festschrift pp. 59–74. *Lecture Notes in Computer Science* 8464. Springer Int. Publ. Switzerland.

Abramsky, S., Barbosa, R. S. & Mansfield, S. (2017). Contextual fraction as a measure of contextuality. *Physical Review Letters* 119, 050504.

Adámek, J., Herrlich, H. & Strecker, G. E. (2004). *Abstract and Concrete Categories: The Joy of Cats.* New York: Wiley. Available at http://katmat.math.uni-bremen.de/acc (Accessed May 26, 2019)

Adams, E. W. (1998). *A Primer of Probabilistic Logic.* Chicago: University of Chicago Press.

Allo, P. (2009). Reasoning about data and information: Abstraction between states and commodities. *Synthese* 167(2), 231–249.

Allwein, G. (2004). A qualititative framework for Shannon Information theories. In *NSPW '04: Proceedings of 2004 Workshop on New Security Paradigms* (Nova Scotia, Canada, September 20-23, 2004). New York: ACM (pp. 23-31).

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

Allwein, G., Moskowitz, I. S. & Chang L.-W. (2004). A new framework for Shannon information theory. Technical Report A801024 Naval Research Laboratory, Washington DC. (17 pp).

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 in Cognitive Science* 7, 166–172.

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

Baker, G. A. Jr. (1958). Formulation of quantum mechanics based on the quasi-probability distribution in phase space. *Physical Review* 109, 2198–2206.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience* 5, 617–629.

Barbosa, R. S., Douce, T., Emeriau, P.-E., Kashefi, E. & Mansfield, S. (2019) Continuous-variable nonlocality and contextuality. *Internation Conference on Quantum Physics and Logic (QPL 2019)*, to appear. arXiv:1905.08267v1[quant-ph]

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

Barr, M. (1991). *-Autonomous categories and linear logic. *Mathematical Structures in Computer Science* 1, 159–178.

Bartlett, S. D., Rudolph, T. & Spekkens, R. W. (2007). Reference frames, superselection rules, and quantum information. *Reviews of Modern Physics* 79, 555–609.

Bartosik, H., Klepp, J., Schmitzer, C., Sponar, S., Cabello, A., Rauch, H. & Hasegawa, Y. (2009). Experimental test of quantum contextuality in neutron interferometry. *Physical Review Letters* 103, 040403.

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 Journal of Formal Logic* 38(4), 488–515.

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

Basieva, I., Cervantes, V. H., Dzhafarov, E. N. & Khrennikov, A. (2019). True contextuality beats directs influences in human decision making. *Journal of Experimental Psychology: General* 148(11), 1925–1937. Available as arXiv:1812.04721

Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics* 1(3), 195–200.

Bell, J. S. (1966). On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* 38, 447–452.

Bohr, N. (1928). The quantum postulate and the recent development of atomic theory. *Nature* 121, 580–590.

Brown, R. & Porter, T. (2006). Category theory: an abstract setting for analogy and comparison. In: *What is Category Theory?* pp. 257–274. Advanced Studies in Mathematics and Logic. Polimetrica Publisher, Italy.

Bubic, A., von Cramon, D. Y. & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience* 4, 25.

Cervantes, V. H. & Dzhafarov, E. N. (2018). Snow Queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. *Decision* 5, 193–204.

Chiribella, G., D'Ariano, G. M. & Perinotti, P. (2016). Quantum from principles. In: Chiribella, G. and Spekkens, R. W. (Eds.) *Quantum Theory: Informational Foundations and Foils* (Fundamental Theories of Physics 181). Dordrecht: Springer (pp. 171–221).

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science, *Behavioral and Brain Science* **36**(3): 1–73.

Clark, A. (2017). How to knit your own Markov blanket: Resisting the second law with metamorphic minds. In (T. Wetzinger and W. Wiese, eds.) *Philosophy and Predictive Processing* 3, 19pp. Frankfurt am Mainz Mind Group. doi:10.15502/9783958573031

Collel, G. & Fauquet, J. (2015). Brain activty and cognition: a connection from thermodynamics and information theory. *Frontiers in Psychology* 6, Article 818, 11 pp. doi:10.3389/fpsyg.2015.00818

Collier, J. (2011). Information, causation and computation. In (G. D. Crnkovic and M. Burgin, eds.) *Information and Computation: Essays on Scientific and Philosophical Foundations of Information and Computation*, pp. 89-105. *World Scientific Series in Information Studies Vol 2.* World Scientific Press. Hackensack, NJ.

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

Culbertson, J. & Sturtz, K. (2013). Bayesian machine learning via category theory. arXiv:1312.1445v1[math.CT]

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

Dennett, D. (1984). Cognitive wheels: The frame problem in artificial intelligence. In: Hookway, C. (Ed.) *Minds, Machines and Evolution.* Cambridge, UK: Cambridge University Press (pp. 129–151).

Dietrich, E. & Fields, C. (1996). The role of the frame problem in Fodor's modularity thesis: A case study of rationalist cognitive science. In: K. M. Ford and Z. W. Pylyshyn (Eds.) The Robot's Dilemma Revisited. Norwood, NJ: Ablex (pp. 9–24).

Dietrich, E. & Fields, C. (2020). Equivalence of the Frame and Halting problems. *Algorithms* 13, 175.

Döring, A. & Frembs, M. (2019). Contextuality and the fundamental theorems of quantum mechanics. arXiv:1910.09591v1[math-ph].

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

Dutta, S., Skowron A. & Chakraborty, M. K. (2019). Information flow in logic for distributed systems: Extending graded consequence. *Information Sciences* 491, 232–250.

Dzhafarov, E. N., Kujala, J. V. & Cervantes, V. H. (2016a). Contextuality-by-default: A brief overview of concepts and terminology. In: H. Atmanspacher, T. Filik and E. Pothos (eds.) *Lecture Notes in Computer Science* 9525, 12-23. (available as arXiv:1504.00530)

Dzhafarov, E. N., Zhang, R. & Kujala, J. (2016b). Is there contextuality in behavioural and social systems? *Philosophical Transactions of The Royal Society A* 374, 20150099.

Dzhafarov, E. N. & Kujala, J. V. (2017a). Contextuality-by-Default 2.0: Systems with binary random variables. In: J. A. Barros, B. Coecke and E. Pothos (eds.) *Lecture Notes in Computer Science* 10106, Springer, Berlin, 16–32.

Dzhafarov, E. N. Cervantes, V. H. & Kujala, J. V. (2017b). Contextuality in canonical systems of random variables. *Philosophical Transactions of The Royal Society A* 375, 20160389 doi:10.1098/rsta.2016.0389

Dzharfarov, E. N. & Kon, M. (2018). On universality of classical probability with contextually labeled random varaibles. *Journal of Mathematical Psychology* 85, 17–24.

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

Ehresmann, A. C. & Gomez-Ramirez, J. (2015). Conciliating neuroscience and phenomenology via category theory. *Progress in Biophysics and Molecular Biology* 119, 347–359.

Emary, C., Lambert, N. & Nori, F. (2014). Leggett-Garg inequalities. *Reports on Progress in Physics* 77, 039501.

Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgement and social cognition. *Annual Review of Psychology* 59, 255–278.

Feynman, R. P. (1987). Negative probability. In: Hiley, B. J. and Peat, F. D. (Eds.) *Quantum Implications: Essays in Honor of David Bohm.* London: Routledge and Kegan Paul (pp. 235–248).

Feynman, R. P., Leighton, R. B. & Sands, M. (1965). *Feynman Lectures on Physics.* New Yorlk: Addison-Wesley.

Fields, C. (2011). Implementation of structure mapping inference by event-file binding and action planning: a model of tool-improvisation analogies. *Psychological Research* 75, 129–142.

Fields, C. (2012). Trajectory recognition as the basis for object individuation: A functional model of object file instantiation and object token encoding. *Frontiers in Psychology: Perception Science* 2, 49.

Fields, C. (2013). Metaphorical motion in mathematical reasoning: further evidence for pre-motor implementation of structure mapping in abstract domains. *Cognitive Processing* 14, 217–229.

Fields, C. (2013). How humans solve the frame problem. *Journal of Experimental and Theoretical Artificial Intelligence* 25, 441–456.

Fields, C. (2018). Some consequences of the thermodynamic cost of system identification. *Entropy* 20, 797.

Fields, C. & Glazebrook, J. F. (2019a). A mosaic of Chu spaces and Channel Theory I: Category-theoretic concepts and tools. *Journal of Experimental and Theoretical Artificial Intelligence* 31(2), 177-213.

Fields, C. & Glazebrook, J. F. (2019b). A mosaic of Chu spaces and Channel Theory II: Applications to object identification and mereological complexity. *Journal of Experimental and Theoretical Artificial Intelligence* 31(2), 237–265.

Fields, C. & Glazebrook, J. F. (2020a). Representing measurement as a thermodynamic symmetry breaking. *Symmetry* 12, 810. doi:10.3390/sym12050810

Fields, C. & Glazebrook, J. F. (2020b). Do Process-1 simulations generate the epistemic feelings that drive Process-2 decision making? *Cognitive Processing*, in press. doi:10.1007/s10339-020-00981-9

Fields, C. & Marcianò, A. (2019). Markov blankets are general physical interaction surfaces. *Physics of Life Reviews* 33, 109–111.

Fields, C. & Marcianò, A. (2019). Sharing nonfungible information requires shared nonfungible information. *Quantum Reports* 1, 252-259.

Fine, A. (1982). Hidden variables, joint probability and the Bell inequalities. *Physical Review Letteers* 48, 291–295.

Flombaum, J. I., Scholl, B. J. & Santos, L. R. (2008). Spatiotemporal priority as a fundamental principle of object persistence. In Hood, B. and Santos, L. (Eds.), *The Origins of Object Knowledge.* New York: Oxford University Press (pp. 135–164).

Fodor, J. A. (1983). *The Modularity of Mind.* Cambridge MA: MIT Press.

Frembs, M. & Döring, A. (2019). No-signalling, contextuality and the arrow of time arXiv:1910.09596v1[quant-ph]

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

Friston, K. J. (2013). Life as we know it. *Journal of The Royal Society Interface* 10, 20130475.

Friston, K. J. & Stephan, K. E. (2007). Free-energy and the brain, *Synthese* 159(3), 417–458 doi:10.1007/s11229-007-9237-y

Friston, K. J. & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of The Royal Society B* 364, 1211–1221.

Friston, K. J., Kilner, J. & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology – Paris* 100, 70–87.

Friston, K. J., Levin, M., Sengupta, B. & Pezzulo, G. (2015). Knowing one's place: a free energy approach to pattern regulation. *Journal of the Royal Society Interface* 12, 20141383.

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T. & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience* 6, 187–214.

Froese, T. & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence* 173, 466–500.

Fuchs, C. A. & Schack, R. (2013). Quantum Bayesian coherence. *Reviews of Modern Physics* 85, 1693–1715.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7(2), 155–170, 1983.

Gentner, D. & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist* 52(1), 45–56.

Goguen, J. A. (1991). A categorical manifesto. *Mathematical Structures in Computer Science* 1, 49–67.

Gratus, J. & Porter, T. (2006) A spatial view of information. *Theoretical Computer Science* 365, 206–215.

Gudder, S. (2019). Contexts in quantum measurement theory. *Foundations of Physics* 49, 647–662.

Hoffman, D. D., Singh, M. & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin and Review* 22, 1480–1506.

Hohwy, J. (2013). The Predictive Mind. Oxford University Press, Oxford, UK.

Jennings, D. & Leifer, M. (2016). No return to classical reality. *Contemporary Physics* 57, 60–82.

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

Khrennikov, A. (2009). *A Contextual Approach to Quantum Formulism.* Berlin: Springer.

Kikuchi, M., Nagasaka, I., Toyoda, S. & Kitamura, S. (2003). A mathematical model of interactions in artifact environment. *Proceedings of Society of Instrument and Control Engineers (SICE) Annual Conference 2003* (pp. 2085–2090). SICE Publishers, Fukui, Japan: Fukui University.

Kirchmair, G., Zähringer, F., Gerritsma, R., Kleinmann, M., Gühne, O., Cabello, A., Blatt, R. & Roos, C. F. (2009). State-independent experimental test of quantum contextuality. *Nature* 460, 494–497.

Knill, D. C. & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neuroscience* 27, 712–719,

Kochen, S. & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17, 59–87.

Kreinovich, V., Liu, G. & Nguyen, H. T. (1999): Chu spaces-A new approach to describing uncertainty in systems. In (J. Ramirez-Angulo, ed.) 42nd Midwest Symposium on Circuits and Systems. IEEE Catalogue No.99CH36356. doi:10.1109/MWSCAS.1999.867296

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

Kuchling, F., Friston, K., Georgiev, G. & Levin, M. (2019) Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems, *Physics of Life Reviews* 33, 88–108.

Lindvall, T. (1992). *Lectures on the Coupling Method*. New York, John Wiley and Sons.

Maloney, L. T. & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research* 50, 2362–2374.

McCarthy, J. & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In: Michie, D. and Meltzer, B. (Eds.) *Machine Intelligence, Vol. 4*. Edinburgh: Edinburgh University Press (pp. 463 502).

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

Mermin, N. D. (1993). Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics* 65, 803–815.

Mermin, N. D. (2019). Making better sense of quantum mechanics. *Reports on Progress in Physics* 82, 012002.

Nguyen, N., Nguyen, H. T., Wu, B. & Kreinovich, V. (2001). Chu spaces: Towards new foundations for fuzzy logic and fuzzy control, with applications to information flow on the world-wide web. *Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII)* 5(3), 149–156.

Nhuy, N. & Van Quang, N. (2001). Chu spaces and conditional probabilities. *Acta Mathematica Vietnamica* 26(2), 241–248.

Old, L. J. & Priss, U. (2001). Metaphor and information flow. In 'Proccedings of the 12th Midwest Artificial Intelligence and Cognitive Science Conference', pp. 99–104.

OPERA Collaboration (2012). Measurement of the neutrino velocity with the OPERA detector in the CNGS beam. *Journal of High-Energy Physics* 10, 093.

Papadopoulos, B. K. & Syropoulos, A (2000). Fuzzy sets and fuzzy relational structures as Chu spaces. *Int. Journal of Uncertainty, Fuzziness and Knowledge Based Systems* 8(4), 471–479.

Parmigiani, G. & Inoue, L. Y. T. (2009). *Decision Theory: Principles and Approaches.* Chichester, UK: Wiley.

Pratt, V. (1995). Chu spaces and their interpretation as concurrent objects. *Lecture Notes in Computer Science* 1000, 392–405.

Pratt, V. (1996). Broadening the denotational semantics of linear logic. In: 'Special Issue on Linear Logic', Tokyo, 1996. *Electronic Notes in Theoretical Computer Science* 3, 155–166. http://www.elsevier.nl/locate/entcs/volume3.html

Pratt, V. (1997). Types as Processes, via Chu spaces. Invited paper. Proceedings 'Express'97: Fourth Workshop on Expressiveness in Concurrency'. Santa Margherita, Italy. September 1997. *Electronic Notes in Theoretical Computer Science* 7 21pp. http://www.elsevier.nl/locate/entcs/volume7.html

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. *Annals of Pure and Applied Logic* 96, 319–333.

Roederer, J. (2005). *Information and its Role in Nature.* Heidelberg: Springer.

Sanborn, A. N. & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Science* 20(12), 883–893.

Scholl, B. J. (2007). Object persistence in philosophy and psychology. *Mind and Language* 22, 563–591.

Schorlemmer, M. (2002). Duality in knowledge sharing. Informatics Research Report EDI-INF-RR-0134. Division of Informatics, University of Edinburgh, 2002. (see also 'International Symposium on Artificial Intelligence and Mathematics' AI&M 2002 (M. Golumbic et al. organizers), Fort Lauderdale, FL USA.

Scully, M. O., Walther, H. & Schleich, W. (1994). Feynman's approach to negative probability in quantum mechanics. *Physical Review A* 49, 1562–1566.

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).

Seth, A. K., Suzuki, K. & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology* 2 Article 395, 16 pp. doi: 10.3389/fpsyg.2011.00395

Shanahan, M. & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition* 98, 157–176.

Shanahan, M. (2016). The frame problem. *Stanford Encyclopedia of Philosophy*, retrieved from http:// plato.stanford.edu/entries/frameproblem/, [Accessed Oct. 17, 2019].

Treisman, A. (2006). Object tokens, binding and visual memory. In Zimmer, H. D., Mecklinger, A. and Lindenberger, U. (Eds.) *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*. Oxford, UK: Oxford University Press (pp. 315–338).

van Benthem, J. (2000). Information transfer across Chu spaces. *Logic Journal of the Interest Group in Pure and Applied Logics (IGPL)* 8(6) (2000), 719–731.

Varacca, D., Völzer, H. & Winskel, G. (2006). Probabilistic event structures and domains. *Theoretical Computer Science* 358 (2-3), 173-191. doi:10.1016/j.tcs.2006.01.015

Winskel, G. (1982) Event structure semantics for CCS and related languages. In: Proceedings of the 9th International Colloquim on Autonoma, Languages and Programming (ICALP), *Lecture Notes in Computer Science* 140, Springer, Berlin, 561–576.

Zhang, G.-Q. & Shen, G. (2006). Approximable concepts, Chu spaces, and information systems. *Theory and Applications of Categories* 17(5), 80–102.