On the information-theoretic structure of distributed measurements

David Balduzzi*

Max Planck Institute for Intelligent Systems, Tübingen, Germany

david.balduzzi@tuebingen.mpg.de

The internal structure of a measuring device, which depends on what its components are and how they are organized, determines how it categorizes its inputs. This paper presents a geometric approach to studying the internal structure of measurements performed by distributed systems such as probabilistic cellular automata. It constructs the quale, a family of sections of a suitably defined presheaf, whose elements correspond to the measurements performed by all subsystems of a distributed system. Using the quale we quantify (i) the information generated by a measurement; (ii) the extent to which a measurement is context-dependent; and (iii) whether a measurement is decomposable into independent submeasurements, which turns out to be equivalent to context-dependence. Finally, we show that only indecomposable measurements are more informative than the sum of their submeasurements.

1 Introduction

Any classical physical system (by which we simply mean any deterministic function) can be taken as a measuring apparatus or input/output device. For example, a thermometer takes inputs from the atmosphere and outputs numbers on a digital display. The thermometer categorizes inputs by temperature and is blind to, say, differences in air pressure.

Classical measurements are formalized as follows:

Definition 1. Given a classical physical system with state space X, a *measuring device* is a function $f : X \to \mathbb{R}$. The output $r \in \mathbb{R}$ is the *reading* and the pre-image $f^{-1}(r) \subset X$ is the *measurement*.

From this point of view a thermometer and a barometer are two functions, $T : X \to \mathbb{R}$ and $B : X \to \mathbb{R}$, mapping the state space X of configurations (positions and momenta) of atmospheric particles to real numbers. When the thermometer outputs $2°$, it specifies that the atmospheric configuration was in the pre-image $T^{-1}(2°)$ which, assuming the thermometer perfectly measures temperature, is *exactly* characterized as atmospheric configurations with temperature $2°$. Similarly, the pre-images generated by the barometer group atmospheric configurations by pressure.

The classical definition of measurement takes a thermometer as a monolithic object described by a single function from atmospheric configurations to real numbers. The internal structure of the thermometer – that is composed of countless atoms and molecules arranged in an extremely specific manner – is swept under the carpet (or, rather, into the function).

This paper investigates the structure of measurements performed by *distributed* systems. We do so by adapting Definition 1 to a large class of systems that contains networks of Boolean functions [18], Conway's game of life [9, 12] and Hopfield networks [3, 15] as special cases.

Our motivation comes from prior work investigating information processing in discrete neural networks [6, 7]. The brain X can be thought of as an enormously complicated measuring device $S \times X \xrightarrow{f} X$

---

mapping sensory states $s \in S$ and prior brain states $x \in X$ to subsequent brain states. Analyzing the functional dependencies implicit in cortical computations reduces to analyzing how the measurements performed by the brain are composed out of submeasurements by subdevices such as individual neurons and neuronal assemblies. The cortex is of particular interest since it seemingly effortlessly integrates diverse contextual data into a unified gestalt that determines behavior. The measurements performed by different neurons appear to interact in such a way that they generate more information jointly than separately. To improve our understanding of how the cortex integrates information we need to a formal language for analyzing how context affects measurements in distributed systems.

As a first step in this direction, we develop methods for analyzing the geometry of measurements performed by functions with overlapping domains. We propose, roughly speaking, to study context-dependence in terms of the geometry of intersecting pre-images. However, since we wish to work with both probabilistic and deterministic systems, things are a bit more complicated.

We sketch the contents of the paper. Section §2 lays the groundwork by introducing the category of stochastic maps `Stoch`. Our goal is to study finite set valued functions and conditional probability distributions on finite sets. However, rather than work with sets, functions and conditional distributions, we prefer to study stochastic maps (Markov matrices) between function spaces on sets. We therefore introduce the faithful functor $\mathscr{V}$ taking functions on sets to Markov matrices:

$$\Big[f : X \to Y\Big] \mapsto \Big[\mathscr{V}f : \mathscr{V}X \to \mathscr{V}Y\Big],$$

where $\mathscr{V}X$ is functions from X to $\mathbb{R}$. Conditional probability distributions $p(y|x)$ can also be represented using stochastic maps.

Working with linear operators instead of set-valued functions is convenient for two reasons. First, it unifies the deterministic and probabilistic cases in a single language. Second, the dual $T^{\natural}$ of a stochastic map T provides a symmetric treatment of functions and their corresponding inverse image functions. Recall the inverse of function $f : X \to Y$ is $f^{-1} : Y \to \underline{2}^X$, which takes values in the *powerset* of X, rather than X itself. Dualizing a stochastic map flips the domain and range of the original map, without introducing any new objects:

$$\Big[f^{-1} : Y \to \underline{2}^X\Big] \text{ corresponds to } \Big[(\mathscr{V}f)^{\natural} : \mathscr{V}Y \to \mathscr{V}X\Big], \tag{1}$$

see Proposition 2.

Section §3 introduces distributed dynamical systems. These extend probabilistic cellular automata by replacing cells (space coordinates) with occasions (spacetime coordinates: cell k at time t). Inspired by [1, 14], we treat distributed systems as collections of stochastic maps between function spaces so that processes (stochastic maps) take center stage, rather than their outputs. Although the setting is abstract, it has the advantage that it is *scalable*: using a coarse-graining procedure introduced in [5] we can analyze distributed systems at any spatiotemporal granularity.

Distributed dynamical systems provide a rich class of toy universes. However, since these toy universes do not contain conscious observers we confront Bell's problem [8]: "What exactly qualifies some physical [system] to play the role of 'measurer'?" In our setting, where we do not have to worry about collapsing wave-functions or the distinction between macroscopic and microscopic processes, the solution is simple: *every* physical system plays the role of measurer. More precisely, we track measurers via the category $\mathsf{Sys_D}$ of subsystems of $\mathbf{D}$. Each subsystem $\mathbf{C}$ is equipped with a mechanism $\mathfrak{m}_{\mathbf{C}}$ which is constructed by gluing together the mechanisms of the occasions in $\mathbf{C}$ and averaging over extrinsic noise.

Measuring devices are typically analyzed by varying their inputs and observing the effect on their outputs. By contrast this paper fixes the output and *varies the device over all its subdevices* to obtain

a family of submeasurements parametrized by all subsystems in $\mathrm{Sys}_{\mathbf{D}}$. The internal structure of the measurement performed by $\mathbf{D}$ is then studied by comparing submeasurements.

We keep track of submeasurements by observing that they are sections of a suitably defined presheaf. Sheaf theory provides a powerful machinery for analyzing relationships between objects and subobjects [19], which we adapt to our setting by introducing the structure presheaf $\mathscr{F}$, a contravariant functor from $\mathrm{Sys}_{\mathbf{D}}$ to the category of measuring devices $\mathrm{Meas}_{\mathbf{D}}$ on $\mathbf{D}$. Importantly, $\mathscr{F}$ is *not* a sheaf: although the gluing axiom holds, uniqueness fails, see Theorem 4. This is because the restriction operator in Meas is (essentially) marginalization, and of course there are infinitely many joint distributions $p(x,y)$ that yield marginals $p(x)$ and $p(y)$.

Section §4 adapts Definition 1 to distributed systems and introduces the simplest quantity associated with a measurement: effective information, which quantifies its precision, see Proposition 5. Crucially, effective information is *context-dependent* – it is computed relative to a baseline which may be completely uninformative (the so-called null system) or provided by a subsystem.

Finally entanglement, introduced in §5, quantifies the obstruction (in bits) to decomposing a measurement into independent submeasurements. It turns out, see discussion after Theorem 10, that entanglement quantifies the extent to which a measurement is context-dependent – the extent to which contextual information provided by one submeasurement is useful in understanding another. Theorem 9 shows that a measurement is more precise than the sum of its submeasurements *only if* entanglement is non-zero. Precision is thus inextricably bound to context-dependence and indecomposability. The failure of unique descent is thus a feature, not a bug, since it provides "elbow room" to build measuring devices that are *not* products of subdevices.

Space constraints prevent us from providing concrete examples; the interested reader can find these in [5–7]. Our running examples are the deterministic set-valued functions

$$f : X \to Y \ \text{ and } \ g : X \times Y \to Z$$

which we use to illustrate the concepts as they are developed.

2   Stochastic maps

Any conditional distribution $p(y|x)$ on finite sets X and Y can be represented as a matrix as follows. Let $\mathscr{V}X = \{\varphi : X \to \mathbb{R}\}$ denote the vector space of real valued functions on X and similarly for Y. $\mathscr{V}X$ is equipped with Dirac basis $\{\delta_x : X \to \mathbb{R} | x \in X\}$, where

$$\delta_x(x') = \begin{cases} 1 & \text{if } x = x' \\ 0 & \text{else.} \end{cases}$$

Given a conditional distribution $p(y|x)$ construct matrix $\mathfrak{m}_p$ with entry $p(y|x)$ in column δ_x and row δ_y. Matrix $\mathfrak{m}_p$ is *stochastic*: it has nonnegative entries and its columns sum to 1. Alternatively, given a stochastic matrix $\mathfrak{m} : \mathscr{V}X \to \mathscr{V}Y$, we can recover the conditional distribution. The Dirac basis induces Euclidean metric

$$\langle \bullet | \bullet \rangle : \mathscr{V}X \otimes \mathscr{V}X \to \mathbb{R} : \Big\langle \sum \alpha_x \delta_x \Big| \sum \beta_x \delta_x \Big\rangle = \sum \alpha_x \beta_x \tag{2}$$

which identifies vector spaces with their duals $\mathscr{V}X \approx (\mathscr{V}X)^*$. Let $p_{\mathfrak{m}}(y|x) := \langle \delta_y | \mathfrak{m}(\delta_x) \rangle$.

Definition 2. The *category of stochastic maps* Stoch has function spaces $\mathscr{V}X$ for objects and stochastic matrices $\mathfrak{m} : \mathscr{V}X \to \mathscr{V}Y$ with respect to Dirac bases for arrows. We identify of $(\mathscr{V}X)^*$ with $\mathscr{V}X$ using the Dirac basis without further comment below.

Definition 3. The *dual* of surjective stochastic map $\mathfrak{m} : \mathscr{V}X \to \mathscr{V}Y$ is the composition $\mathfrak{m}^\natural := \mathscr{V}Y \xrightarrow{\mathfrak{m}^* \circ ren} \mathscr{V}X$, where *ren* is the unique map making diagram

$$
\begin{array}{ccc}
(\mathscr{V}Y)^* & \xrightarrow{\ \mathfrak{m}^*\ } & (\mathscr{V}X)^* \\
{\scriptstyle ren}\uparrow & & \downarrow{\scriptstyle \omega_X} \\
(\mathscr{V}Y)^* & \xrightarrow[\ \omega_Y\ ]{} & \mathbb{R}
\end{array}
$$

commute. Precomposing $\mathfrak{m}^*$ with *ren* renormalizes[1] its columns to sum to 1. The stochastic dual of a stochastic transform is stochastic; further, if $\mathfrak{m}$ is stochastic then $(\mathfrak{m}^\natural)^\natural = \mathfrak{m}$.

Category $\mathtt{Stoch}$ is described in terms of braid-like generators and relations in [11]. A more general, but also more complicated, category of conditional distributions was introduced by Giry [13], see [20].

Example 1 (deterministic functions). Let $\mathtt{FSet}$ be the category of finite sets. Define faithful functor $\mathscr{V} : \mathtt{FSet} \to \mathtt{Stoch}$ taking set X to $\mathscr{V}X$ and function $f : X \to Y$ to stochastic map $\mathscr{V}f : \mathscr{V}X \to \mathscr{V}Y : \delta_x \mapsto \delta_{f(x)}$. It is easy to see that $\mathscr{V}(X \times Y) = \mathscr{V}X \otimes \mathscr{V}Y$ and $\mathscr{V}(X \cup Y) = \mathscr{V}X \times \mathscr{V}Y$.

We introduce special notation for commonly used functions:

- *Set inclusion.* For any inclusion $i : X \hookrightarrow Y$ of sets, let $\iota := \mathscr{V}i : \mathscr{V}X \to \mathscr{V}Y$ denote the corresponding stochastic map. Two important examples are

  - *Point inclusion.* Given $x \in X$ define $\iota_x : \mathbb{R} \to \mathscr{V}X : 1 \mapsto \delta_x$.
  - *Diagonal map.* Inclusion $\Delta : X \hookrightarrow X \times X : x \mapsto (x,x)$ induces $\iota_\Delta : \mathscr{V}X \to \mathscr{V}X \otimes \mathscr{V}X : \delta_x \mapsto \delta_x \otimes \delta_x$.

- *Terminal map.* Let $\omega_X : \mathscr{V}X \to \mathbb{R} : \delta_x \mapsto 1$ denote the terminal map induced by $X \to \{\bullet\}$.

- *Projection.* Let $\pi_{XY,X} : \mathscr{V}X \otimes \mathscr{V}Y \to \mathscr{V}X : \delta_x \otimes \delta_y \mapsto \delta_x$ denote the projection induced by $pr_{X \times Y, X} : X \times Y \to X : (x,y) \mapsto x$.

Proposition 1 (dual is Bayes over uniform distribution)**.** *The dual of a stochastic map applies Bayes rule to compute the posterior distribution* $\langle \mathfrak{m}^\natural(\delta_y)|\delta_x \rangle = p_\mathfrak{m}(x|y)$ *using the uniform probability distribution.*

Proof: The uniform distribution is the dual $\omega_X^\natural : \mathbb{R} \to \mathscr{V}X : 1 \mapsto \frac{1}{|X|} \sum_x \delta_x$ of the terminal map $\omega_X : \mathscr{V}X \to \mathbb{R}$. It assigns equal probability $p_{\omega^\natural}(x) = \frac{1}{|X|}$ to all of X's elements, and can be characterized as the maximally uninformative distribution [16]. Let $\mathfrak{m} : \mathscr{V}X \to \mathscr{V}Y$. The normalized transpose is

$$
\mathfrak{m}^\natural(\delta_y) = \sum_x \frac{p_\mathfrak{m}(y|x)}{\sum_{x'} p_\mathfrak{m}(y|x')} \delta_x = \sum_x \frac{p_\mathfrak{m}(y|x) \cdot p_{\omega^\natural}(x)}{\sum_{x'} p_\mathfrak{m}(y|x') p_{\omega^\natural}(x')} \delta_x = \sum_x p_\mathfrak{m}(x|y) \cdot \delta_x. \ \blacksquare
$$

Remark 1. Note that $p_\mathfrak{m}(x|y) := \langle \mathfrak{m}^\natural(\delta_y)|\delta_x \rangle \neq \langle \delta_y|\mathfrak{m}(\delta_x) \rangle =: p_\mathfrak{m}(y|x)$. Dirac's bra-ket notation must be used with care since stochastic matrices are not necessarily symmetric [10].

Corollary 2 (preimages)**.** *The dual* $(\mathscr{V}f)^\natural : \mathscr{V}Y \to \mathscr{V}X$ *of stochastic map* $\mathscr{V}f : \mathscr{V}X \to \mathscr{V}Y$ *is conditional distribution*

$$
p_{\mathscr{V}f}(x|y) = \begin{cases} \frac{1}{|f^{-1}(y)|} & \text{if } f(x) = y \\ 0 & \text{else.} \end{cases} \tag{3}
$$

---

[1] If $\mathfrak{m}$ is not surjective, i.e. if one of the rows has all zero entries, then the renormalization is not well-defined.

Proof: By the proof of Proposition 1

$$(\mathscr{V}f)^{\natural}(\delta_y) = \frac{1}{|f^{-1}(y)|} \sum_{\{x|f(x)=y\}} \delta_x. \ \blacksquare$$

The support of $p_{\mathscr{V}f}(X|y)$ is $f^{-1}(y)$. Elements in the support are assigned equal probability, thereby treating them as an undifferentiated list. Dual $(\mathscr{V}f)^{\natural}$ thus generalizes the inverse image $f^{-1} : Y \to \underline{2}^X$. Conveniently however, the dual $(\mathscr{V}X)^{\natural}$ simply flips the domain and range of $\mathscr{V}f$, whereas the inverse image maps to powerset $\underline{2}^X$, an entirely new object.

Corollary 3 (marginalization with respect to uniform distribution). *Precomposing $\mathscr{V}X \otimes \mathscr{V}Y \xrightarrow{\mathrm{m}} \mathscr{V}Z$ with the dual $\pi_X^{\natural}$ to $\mathscr{V}X \otimes \mathscr{V}Y \xrightarrow{\pi_X} \mathscr{V}X$ marginalizes $p_{\mathrm{m}}(z|x,y)$ over the uniform distribution on Y.*

Proof: By Corollary 2 we have $\pi_X^{\natural} : \mathscr{V}X \to \mathscr{V}X \otimes \mathscr{V}Y : \delta_y \mapsto \frac{1}{|Y|} \sum_{y \in Y} \delta_x \otimes \delta_y$. It follows immediately that

$$p_{\mathrm{m}\circ\pi_X^{\natural}}(z|x) = \frac{1}{|Y|} \sum_{y \in Y} p_{\mathrm{m}}(z|x,y). \ \blacksquare$$

Precomposing with $\pi_X^{\natural}$ treats inputs from Y as extrinsic noise. Although duals can be defined so that they implement Bayes' rule with respect to other probability distributions, this paper restricts attention to the simplest possible renormalization of columns, Definition 2. The uniform distribution is convenient since it uses minimal prior knowledge (it depends only on the number of elements in the set) to generalize pre-images to the stochastic case, Proposition 2.

3   Distributed dynamical systems

Probabilistic cellular automata provide useful toy models of a wide range of physical and biological systems. A cellular automaton consists of a collection of cells, each equipped with a mechanism whose output depends on the prior outputs of its neighbors. Two important important examples are

Example 2 (Conway's game of life). The cellular automaton is a grid of deterministic cells with outputs $\{0,1\}$. A cell outputs 1 at time t iff: (i) three of its neighbors outputted 1s at time $t-1$ or (ii) it and two neighbors outputted 1s at $t-1$. Remarkably, a sufficiently large game of life grid can implement any deterministic computation [9].

Example 3 (Hopfield networks). These are probabilistic cellular automata [3, 15], again with outputs $\{0,1\}$. Cell n_k fires with probability proportional to

$$p(n_{k,t} = 1 | n_{\bullet,t-1}) \propto \exp\left[\frac{1}{T} \sum_{j \to k} \alpha_{jk} \cdot n_{j,t-1}\right].$$

Temperature T controls network stochasticity. Attractors $\{\xi^1, \ldots, \xi^N\}$ are embedded into a network by setting the connectivity matrix as $\alpha_{jk} = \sum_{\mu=1}^{N} (2\xi_j^{\mu} - 1)(2\xi_k^{\mu} - 1)$.

It is useful to take a finer perspective on cellular automata by decomposing them into spacetime coordinates or occasions [5]. An occasion $v_l = n_{i,t}$ is a cell n_i at a time point t. Two occasions are linked $v_k \to v_l$ if there is a connection from v_k's cell to v_l's (because they are neighbors or the same cell) and their time coordinates are $t-1$ and t respectively for some t, so occasions form a directed graph. More generally:

Definition 4. A *distributed dynamical system* **D** consists of the following data:

D1. *Directed graph.* A graph $G_{\mathbf{D}} = (V_{\mathbf{D}}, E_{\mathbf{D}})$ with a finite set of vertices or occasions $V_{\mathbf{D}} = \{v_1 \ldots v_n\}$ and edges $E_{\mathbf{D}} \subset V_{\mathbf{D}} \times V_{\mathbf{D}}$.

D2. *Alphabets.* Each vertex $v_l \in V_{\mathbf{D}}$ has finite alphabet A_l of outputs and finite alphabet $S_l := \prod_{k \in src(l)} A_k$ of inputs, where $src(l) = \{v_k | v_k \to v_l\}$.

D3. *Mechanisms.* Each vertex v_l is equipped with stochastic map $\mathfrak{m}_l : \mathscr{V} S_l \to \mathscr{V} A_l$.
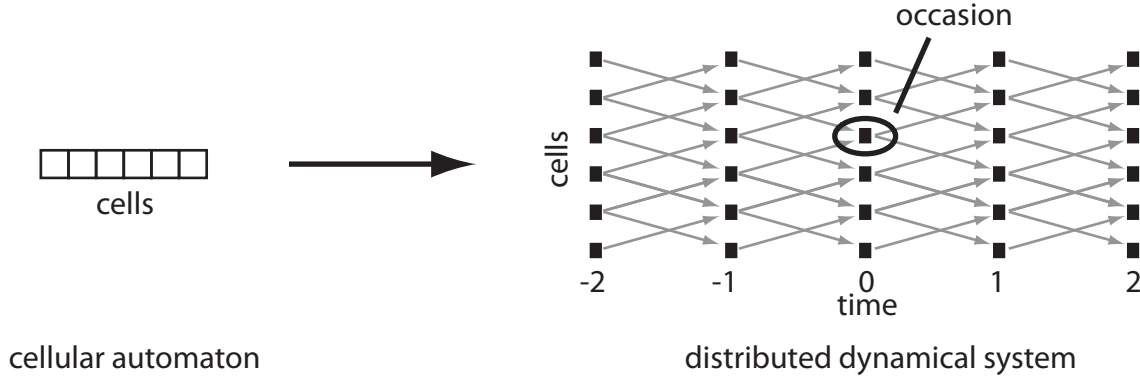

Figure 1: Mapping a cellular automaton to a distributed dynamical system.

Taking any cellular automaton over a finite time interval $[t_\alpha, t_\omega]$ initializing the mechanisms at time t_α with fixed values (initial conditions) or probability distributions (noise sources) yields a distributed dynamical system, see Fig. 1. Each cell of the original automaton corresponds to a series of occasions in the distributed dynamical system, one per time step.

Cells with memory – i.e. whose outputs depend on their neighbors outputs over multiple time steps – receive inputs from occasions more than one time step in the past. If a cell's mechanism changes (learns) over time then different mechanisms are assigned to the cell's occasions at different time points.

The sections below investigate the compositional structure of measurements: how they are built out of submeasurements. Technology for tracking subsystems and submeasurements is therefore necessary. We introduce two closely related categories:

Definition 5. The *category of subsystems* $\mathtt{Sys}_{\mathbf{D}}$ of $\mathbf{D}$ is a Boolean lattice with objects given by sets of ordered pairs of vertices $\mathbf{C} \in 2^{V_{\mathbf{D}} \times V_{\mathbf{D}}}$ and arrows given by inclusions $i_{12} : \mathbf{C}_1 \hookrightarrow \mathbf{C}_2$. The initial and terminal objects are $\perp_{\mathbf{D}} = \emptyset$ and $\top_{\mathbf{D}} = V_{\mathbf{D}} \times V_{\mathbf{D}}$.

Remark 2. Subsystems are defined as ordered pairs of vertices, rather than subgraphs of the directed graph of $\mathbf{D}$. Pairs of occasions that are not connected by edges are *ineffective*; they do not contribute to the information-processing performed by the system. We include them in the formalism precisely to make their lack of contribution explicit, see Remark 3.

Let $src(\mathbf{C}) = \{v_k | (v_k, v_l) \in \mathbf{C}\}$ and similarly for $trg(\mathbf{C})$. Set the input alphabet of $\mathbf{C}$ as the product of the output alphabets of its source occasions $S^{\mathbf{C}} = \prod_{src(\mathbf{C})} A_k$ and similarly the output alphabet of $\mathbf{C}$ as the product of the output alphabets of its target occasions $A^{\mathbf{C}} = \prod_{trg(\mathbf{C})} A_l$.

Definition 6. The *category of measuring devices* $\mathtt{Meas}_{\mathbf{D}}$ on $\mathbf{D}$ has objects $\mathrm{Hom}_{\mathtt{Stoch}}(\mathscr{V} A^{\mathbf{C}}, \mathscr{V} S^{\mathbf{C}})$ for

$\mathbf{C} \in \underline{2}^{V_{\mathbf{D}} \times V_{\mathbf{D}}}$. For $\mathbf{C}_1 \hookrightarrow \mathbf{C}_2$ define arrow

$$r_{21} : \mathrm{Hom}\left(\mathscr{V}A^{\mathbf{C}_2}, \mathscr{V}S^{\mathbf{C}_2}\right) \to \mathrm{Hom}\left(\mathscr{V}A^{\mathbf{C}_1}, \mathscr{V}S^{\mathbf{C}_1}\right)$$

$$\left[\mathscr{V}A^{\mathbf{C}_2} \xrightarrow{T} \mathscr{V}S^{\mathbf{C}_2}\right] \mapsto \left[\mathscr{V}A^{\mathbf{C}_1} \xrightarrow{\pi_A^\natural} \mathscr{V}A^{\mathbf{C}_2} \xrightarrow{T} \mathscr{V}S^{\mathbf{C}_2} \xrightarrow{\pi_S} \mathscr{V}S^{\mathbf{C}_1}\right],$$

where π_A and π_S are shorthands for projections as in Definition 1.

The reason for naming $\mathtt{Meas_D}$ the category of measuring devices will become clear in §4 below. The two categories are connected by contravariant functor $\mathscr{F}$:

Theorem 4 (structure presheaf). *The* structure presheaf $\mathscr{F}$ *taking*

$$\mathscr{F}_{\mathbf{D}} : \mathtt{Sys}_{\mathbf{D}}^{\mathrm{op}} \to \mathtt{Meas_D} : \mathbf{C} \mapsto \mathrm{Hom}\left(\mathscr{V}A^{\mathbf{C}}, \mathscr{V}S^{\mathbf{C}}\right) \text{ and } i_{12} \mapsto r_{21}$$

satisfies the gluing axiom but has non-unique descent.

Proof: Functor $\mathscr{F}$ is trivially a presheaf since it is contravariant. It is an *interesting* presheaf because the gluing axiom holds.

For gluing we need to show that for any collection $\{\mathbf{C}_j\}_{j=1}^l$ of subsystems and sections $\mathfrak{m}_j^\natural \in \mathscr{F}_{\mathbf{D}}(\mathbf{C}_j)$ such that $r_{j,ji}(\mathfrak{m}_j^\natural) = r_{i,ji}(\mathfrak{m}_i^\natural)$ for all i, j there exists section $\mathfrak{m}^\natural \in \mathscr{F}_{\mathbf{D}}\left(\bigcup_{j=1}^l \mathbf{C}_j\right)$ such that $r_i(\mathfrak{m}^\natural) = \mathfrak{m}_i^\natural$ for all i. This reduces to finding a conditional distribution that causes diagram

$$
\begin{array}{ccc}
? & \longrightarrow & \mathfrak{m}_i \\
\downarrow & & \downarrow \\
\mathfrak{m}_j & \longrightarrow & \mathfrak{m}_{ij}
\end{array}
$$

in $\mathtt{Meas_D}$ to commute. The vertices are conditional distributions and the arrows are marginalizations, so rewrite as

$$
\begin{array}{ccc}
? & \longrightarrow & p(x,y|u,w) \\
\downarrow & & \downarrow \\
p(x,z|v,w) & \longrightarrow & p(x|w),
\end{array}
$$

where $p(x|w) = \sum_{v,z} p(x,z|v,w) p^{maxH}(v)$ and similarly for the vertical arrow. It is easy to see that

$$p(x,y,z|u,v,w) := \frac{p(x,y|u,w)p(x,z|v,w)}{p(x|w)}$$

satisfies the requirement.

For $\mathscr{F}$ to be a sheaf it would also have to satisfy unique descent: the section satisfying the gluing axiom must not only *exist* for any collection $\{\mathbf{C}_j\}_{j=1}^l$ with compatible restrictions but must also be *unique*. Descent in $\mathscr{F}$ is not unique because there are many distributions satisfying the requirement above: strictly speaking r is a marginalization operator rather than restriction. For example, there are many distributions $p(y,z)$ that marginalize to give $p(y)$ and $p(z)$ besides the product distribution $p(y)p(z)$. ∎

The structure presheaf $\mathscr{F}$ depends on the graph structure and alphabets; mechanisms play no role. We now construct a family of sections of $\mathscr{F}$ using the mechanisms of $\mathbf{D}$'s occasions. Specifically, given a

subsystem $\mathbf{C} \in \mathrm{Sys}_{\mathbf{D}}$, we show how to glue its occasions' mechanisms together to form joint mechanism $\mathfrak{m}_{\mathbf{C}}$. The mechanism $\mathfrak{m}_{\mathbf{D}} = \mathfrak{m}_{\top}$ of the entire system $\mathbf{D}$ is recovered as a special case.

In general, subsystem $\mathbf{C}$ is not isolated: it receives inputs along edges contained in $\mathbf{D}$ but *not* in $\mathbf{C}$. Inputs along these edges cannot be assigned a fixed value since in general there is no preferred element of A_l. They also cannot be ignored since $\mathfrak{m}_l$ is defined as receiving inputs from all its sources. Nevertheless, the mechanism of $\mathbf{C}$ should depend on $\mathbf{C}$ alone. We therefore treat edges not in $\mathbf{C}$ as sources of extrinsic noise by marginalizing with respect to the uniform distribution as in Corollary 3.

For each vertex $v_l \in trg(\mathbf{C})$ let $S_l^{\mathbf{C}} = \prod_{k \in src(l) \cap src(C)} A_k$. We then have projection $\pi_l : \mathscr{V} S_l \to \mathscr{V} S_l^{\mathbf{C}}$. Define

$$\mathfrak{m}_l^{\mathbf{C}} := \left[\mathscr{V} S_l^{\mathbf{C}} \xrightarrow{\pi_l^{\natural}} \mathscr{V} S_l \xrightarrow{\mathfrak{m}_l} \mathscr{V} A_l \right]. \tag{4}$$

It follows immediately that $\mathbf{C}$ is itself a distributed dynamical system defined by its graph, whose alphabets are inherited from $\mathbf{D}$ and whose mechanisms are constructed by marginalizing.

Next, we tensor the mechanisms of individual occasions and glue them together using the diagonal map $\Delta : S^{\mathbf{C}} \to \prod_{v_l \in trg(\mathbf{C})} S_l^{\mathbf{C}}$. The diagonal map used here[2] generalizes $X \xrightarrow{\Delta} X \times X$ and removes redundancies in $\prod_l S_l^{\mathbf{C}}$, which may, for example, include the same source alphabets many times in different factors.

Let mechanism $\mathfrak{m}_{\mathbf{C}}$ be

$$\mathfrak{m}_{\mathbf{C}} := \left[\mathscr{V} S^{\mathbf{C}} \xrightarrow{\iota_{\Delta}} \bigotimes_{v_l \in trg(\mathbf{C})} \mathscr{V} S_l^{\mathbf{C}} \xrightarrow{\otimes_{v_l \in trg(\mathbf{C})} \mathfrak{m}_l^{\mathbf{C}}} \mathscr{V} A^{\mathbf{C}} \right]. \tag{5}$$

The dual of $\mathfrak{m}_{\mathbf{C}}$ is

$$\mathfrak{m}_{\mathbf{C}}^{\natural} := \left[\mathscr{V} A^{\mathbf{C}} \to \mathscr{V} S^{\mathbf{C}} \right]. \tag{6}$$

Finally, we find that we have constructed a family of sections of $\mathscr{F}$:

Definition 7. The *quale* $\mathbf{q}_{\mathbf{D}}$ is the family of sections of $\mathscr{F}$ constructed in Eqs. (4), (5) and (6)

$$\mathbf{q}_{\mathbf{D}} := \left\{ \mathfrak{m}_{\mathbf{C}}^{\natural} \in \mathscr{F}(\mathbf{C}) = \mathrm{Hom}\left(\mathscr{V} A^{\mathbf{C}}, \mathscr{V} S^{\mathbf{C}} \right) \Big| \mathbf{C} \in \mathrm{Sys}_{\mathbf{D}} \right\}.$$

The construction used to glue together the mechanism of the entire system can also be used to construct the mechanism of any subsystem, which provides a window – the quale – into the compositional structure of distributed processes.

4 Measurement

This section adapts Definition 1 to distributed stochastic systems. The first step is to replace elements of state space X with stochastic maps $d_{in} : \mathbb{R} \to \mathscr{V} S^{\mathbf{D}}$, or equivalently probability distributions on $S^{\mathbf{D}}$, which are the system's inputs. Individual elements of $S^{\mathbf{D}}$ correspond to Dirac distributions.

Second, replace function $f : X \to \mathbb{R}$ with mechanism $\mathfrak{m}_{\mathbf{D}} : \mathscr{V} S^{\mathbf{D}} \to \mathscr{V} A^{\mathbf{D}}$. Since we are interested in the compositional structure of measurements we also consider submechanisms $\mathfrak{m}_{\mathbf{C}}$. However, comparing mechanisms requires that they have the same domain and range, so we extend $\mathfrak{m}_{\mathbf{C}}$ to the entire system as follows

$$\mathfrak{m}_{\mathbf{C}} = \mathscr{V} S^{\mathbf{D}} \xrightarrow{\pi} \mathscr{V} S^{\mathbf{C}} \xrightarrow{\mathfrak{m}_{\mathbf{C}}} \mathscr{V} A^{\mathbf{C}} \xrightarrow{\pi^{\natural}} \mathscr{V} A^{\mathbf{D}}. \tag{7}$$

---

[2]which is surjective in the sense that all rows contain non-zero entries

We refer to the extension as $\mathfrak{m}_{\mathbf{C}}$ by abuse of notation. We extend mechanisms implicitly whenever necessary without further comment. Extending mechanisms in this way maps the quale into a cloud of points in $\mathrm{Hom}(\mathscr{V}A^{\mathbf{D}}, \mathscr{V}S^{\mathbf{D}})$ labeled by objects in $\mathrm{Sys}_{\mathbf{D}}$.

In the special case of the initial object $\perp_{\mathbf{D}}$, define

$$\mathfrak{m}_{\perp} = \mathscr{V}S^{\mathbf{D}} \xrightarrow{\omega} \mathbb{R} \xrightarrow{\omega^{\natural}} \mathscr{V}A^{\mathbf{D}}.$$

Remark 3. Subsystems differing by non-existent edges (Remark 2) are mapped to the same mechanism by this construction, thus making the fact that the edges do not exist explicit within the formalism.

Composing an input with a submechanism yields an output $d_{out} := \mathfrak{m}_{\mathbf{C}} \circ d_{in} : \mathbb{R} \to \mathscr{V}A^{\mathbf{D}}$, which is a probability distribution on $A^{\mathbf{D}}$. We are now in a position to define

Definition 8. A *measuring device* is the dual $\mathfrak{m}_{\mathbf{C}}^{\natural}$ to the mechanism of a subsystem. An *output* is a stochastic map $d_{out} : \mathbb{R} \to \mathscr{V}A^{\mathbf{D}}$. A *measurement* is a composition $\mathfrak{m}_{\mathbf{C}}^{\natural} \circ d_{out} : \mathbb{R} \to \mathscr{V}S^{\mathbf{D}}$.

Recall that stochastic maps of the form $\mathbb{R} \to \mathscr{V}X$ correspond to probability distributions on X. Outputs as defined above are thus probability distributions on $A^{\mathbf{D}}$, the output alphabet of $\mathbf{D}$. Individual elements of $A^{\mathbf{D}}$ are recovered as Dirac vectors: $\mathbb{R} \xrightarrow{\delta_a} \mathscr{V}A^{\mathbf{D}}$.

Definition 9. The *effective information* generated by $\mathbf{C}_1$ in the *context* of subsystem $\mathbf{C}_2 \subset \mathbf{C}_1$ is

$$ei(\mathfrak{m}_{\mathbf{C}_2} \to \mathfrak{m}_{\mathbf{C}_1}, d_{out}) := H\left[\mathfrak{m}_{\mathbf{C}_1}^{\natural} \circ d_{out} \,\middle\|\, \mathfrak{m}_{\mathbf{C}_2}^{\natural} \circ d_{out}\right]. \tag{8}$$

The *null context*, corresponding to the empty subsystem $\perp = \emptyset \subset V_{\mathbf{D}} \times V_{\mathbf{D}}$, is a special case where $\mathfrak{m}_{\mathbf{C}}^{\natural} \circ d_{out}$ is replaced by the uniform distribution $\omega_{\mathbf{D}}^{\natural}$ on $S^{\mathbf{D}}$. To simplify notation define

$$ei(\mathfrak{m}_{\mathbf{C}}, d_{out}) := ei(\mathfrak{m}_{\perp} \to \mathfrak{m}_{\mathbf{C}}, d_{out}).$$

Here, $H[p\|q] = \sum_i p_i \log_2 \frac{p_i}{q_i}$ is the Kullback-Leibler divergence or relative entropy [17]. Eq. (8) expands as

$$ei(\mathfrak{m}_{\mathbf{C}_2} \to \mathfrak{m}_{\mathbf{C}_1}, d_{out}) = \sum_{s \in S^{\mathbf{D}}} \left\langle \mathfrak{m}_{\mathbf{C}_1}^{\natural} \circ d_{out} \,\middle|\, \delta_s \right\rangle \cdot \log_2 \frac{\left\langle \mathfrak{m}_{\mathbf{C}_1}^{\natural} \circ d_{out} \,\middle|\, \delta_s \right\rangle}{\left\langle \mathfrak{m}_{\mathbf{C}_2}^{\natural} \circ d_{out} \,\middle|\, \delta_s \right\rangle}. \tag{9}$$

When $d_{out} = \delta_a$ for some $a \in A^{\mathbf{D}}$ we have

$$ei(\mathfrak{m}_{\mathbf{C}_2} \to \mathfrak{m}_{\mathbf{C}_1}, \delta_a) = \sum_{s \in S^{\mathbf{D}}} p_{\mathfrak{m}_{\mathbf{C}_1}}(s|a) \cdot \log_2 \frac{p_{\mathfrak{m}_{\mathbf{C}_1}}(s|a)}{p_{\mathfrak{m}_{\mathbf{C}_2}}(s|a)}. \tag{10}$$

Definition 8 requires some unpacking. To relate it to the classical notion of measurement, Definition 1, we consider system $\mathbf{D} = \left\{ v_X \xrightarrow{f} v_Y \right\}$ where the alphabets of v_X and v_Y are the sets $A_{v_X} = X$ and $A_{v_Y} = Y$ respectively, and the mechanism of v_Y is $\mathfrak{m}_Y = \mathscr{V}f$. In other words, system $\mathbf{D}$ corresponds to a single deterministic function $f : X \to Y$.

Proposition 5 (classical measurement). *The measurement* $(\mathscr{V}f)^{\natural} \circ \delta_y$ *performed when deterministic function* $f : X \to Y$ *outputs* y *is equivalent to the preimage* $f^{-1}(y)$. *Effective information is* $ei(\mathscr{V}f, \delta_y) = \log_2 \frac{|X|}{|f^{-1}(y)|}$.

Proof: By Corollary 2 measurement $(\mathscr{V}f)^{\natural} \circ \delta_y$ is conditional distribution

$$p_{\mathscr{V}f}(x|y) = \begin{cases} \frac{1}{|f^{-1}(y)|} & \text{if } f(x) = y \\ 0 & \text{else.} \end{cases}$$

which generalizes the preimage. Effective information follows immediately. ∎

Effective information can be interpreted as quantifying a measurement's precision. It is high if few inputs cause f to output y out of many – i.e. $f^{-1}(y)$ has few elements relative to $|X|$ – and conversely is low if many inputs cause f to output y – i.e. if the output is relatively insensitive to changes in the input. Precise measurements say a lot about what the input could have been and conversely for vague measurements with low *ei*.

The point of this paper is to develop techniques for studying measurements constructed out of two or more functions. We therefore present computations for the simplest case, distributed system $X \times Y \xrightarrow{g} Z$, in considerable detail. Let **D** be the graph

$$v_X \searrow \qquad \swarrow v_Y$$
$$v_Z$$
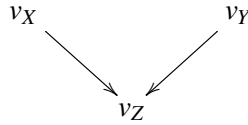
with obvious assignments of alphabets and the mechanism of v_Z as $\mathfrak{m}_Z = \mathscr{V}g$. To make the formulas more readable let $\mathfrak{m}_{XY} = \mathscr{V}g$, $\mathfrak{m}_{X\bullet} = \mathscr{V}g \circ \pi_{XY,X}^{\natural}$ and $\mathfrak{m}_{\bullet Y} = \mathscr{V}g \circ \pi_{XY,Y}^{\natural}$. We then obtain lattice

$$\top = \mathfrak{m}_{XY} \qquad\qquad (11)$$

$$ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}) \nearrow \qquad \nwarrow ei(\mathfrak{m}_{\bullet Y} \to \mathfrak{m}_{XY})$$

$$\mathfrak{m}_{X\bullet} \qquad ei(\mathfrak{m}_{XY}) \qquad \mathfrak{m}_{\bullet Y}$$

$$ei(\mathfrak{m}_{X\bullet}) \searrow \qquad \swarrow ei(\mathfrak{m}_{\bullet Y})$$

$$\mathfrak{m}_{\perp}$$

The remainder of this section and most of the next analyzes measurements in the lattice.

Proposition 6 (partial measurement)**.** *The measurement performed on X when $g : X \times Y \to Z$ outputs z, treating Y as extrinsic noise, is conditional distribution*

$$p(x|z) = \begin{cases} \frac{|g_{x\times Y}^{-1}(z)|}{|g^{-1}(z)|} & \text{if } g(x,y) = z \text{ for some } y \in Y \\ 0 & \text{else,} \end{cases} \qquad (12)$$

where $g_{x\times Y}^{-1}(z) := pr_Y(g^{-1}(z) \cap \{x\} \times Y)$. The effective information generated by the partial measurement is

$$ei(\mathfrak{m}_{X\bullet}^{\natural}, \delta_z) = \log_2 |X| + \sum_{x \in X} p(x|z) \cdot \log_2 p(x|z). \qquad (13)$$

Proof: Treating Y as a source of extrinsic noise yields $\mathscr{V}X \xrightarrow{\pi^{\natural}} \mathscr{V}X \otimes \mathscr{V}Y \xrightarrow{\mathscr{V}g} \mathscr{V}Z$ which takes $\delta_x \mapsto \frac{1}{|Y|} \sum_{y \in Y} \delta_{g(x,y)}$. The dual is

$$\mathfrak{m}_{X\bullet}^{\natural} = \pi_{XY,X} \circ (\mathscr{V}g)^{\natural} : \delta_z \mapsto \sum_{x \in X} \frac{|g_{x\times Y}^{-1}(z)|}{|g^{-1}(z)|} \cdot \delta_x.$$

The computation of effective information follows immediately. ∎

A partial measurement is precise if the preimage $g^{-1}(z)$ has small or empty intersection with $\{x\} \times Y$ for most x, and large intersection for few x.

Propositions 5 and 6 compute effective information of a measurement relative to the null context provided by complete ignorance (the uniform distribution). We can also compute the effective information generated by a measurement in the context of a submeasurement:

Proposition 7 (relative measurement). *The information generated by measurement $X \times Y \xrightarrow{g} Z$ in the context of the partial measurement where Y is unobserved noise, is*

$$ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}, \delta_z) = \sum_{x \in X} \frac{g_{x \times Y}^{-1}(z)}{g^{-1}(z)} \log_2 \frac{|Y|}{g_{x \times Y}^{-1}(z)}. \tag{14}$$

Proof: Applying Propositions 5 and 6 obtains

$$ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}, \delta_z) = \sum_{(x,y) \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \log_2 \left[\frac{1}{|g^{-1}(z)|} \cdot \frac{|g^{-1}(z)| \cdot |Y|}{|g_{x \times Y}^{-1}(z)|} \right]$$

which simplifies to the desired expression. ∎

To interpret the result decompose $X \times Y \xrightarrow{g} Z$ into a family of functions $\mathscr{R} = \left\{ Y \xrightarrow{g_{x \times Y}} Z \,\middle|\, x \in X \right\}$ labeled by elements of X, where $g_{x \times Y}(y) := g(x, y)$. The precision of the measurement performed by $g_{x \times Y}$ is $\log_2 \frac{|Y|}{g_{x \times Y}^{-1}(z)}$. It follows that the precision of the relative measurement, Eq. (14), is the expected precision of the measurements performed by family $\mathscr{R}$ taken with respect to the probability distribution $p(x|z) = \frac{g_{x \times Y}^{-1}(z)}{g^{-1}(z)}$ generated by the noisy measurement.

In the special case of $g : X \times Y \to Z$ relative precision is simply the difference of the precision of the larger and smaller subsystems:

Corollary 8 (comparing measurements).

$$ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}, \delta_z) = ei(\mathfrak{m}_{XY}, \delta_z) - ei(\mathfrak{m}_{X\bullet}, \delta_z)$$

Proof: Applying Propositions 5, 6, 7 and simplifying obtains

$$ei(\mathfrak{m}_{XY}, \delta_z) - ei(\mathfrak{m}_{X\bullet}, \delta_z) = \log_2 \frac{|X| \cdot |Y|}{|g^{-1}(z)|} - \sum_x \frac{|g_{x \times Y}^{-1}(z)|}{|g^{-1}(z)|} \log_2 \frac{|X| \cdot |g_{x \times Y}^{-1}(z)|}{|g^{-1}(z)|}$$

$$= \log_2 \frac{|Y|}{|g^{-1}(z)|} + \sum_{(x,y) \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \log_2 \frac{|g^{-1}(z)|}{|g_{x \times Y}^{-1}(z)|}$$

$$= ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}, \delta_z). \ \blacksquare$$

5 Entanglement

The proof of Theorem 4 showed the structure presheaf has non-unique descent, reflecting the fact that measuring devices do not necessarily reduce to products of subdevices. Similarly, as we will see, measurements do not in general decompose into independent submeasurements. Entanglement, γ, quantifies how far a measurement diverges in bits from the product of its submeasurements. It turns out that $\gamma > 0$ is necessary for a system to generate more information than the sum of its components: non-unique descent

thus provides "room at the top" to build systems that perform more precise measurements collectively than the sum of their components.

Entanglement has no direct relation to quantum entanglement. The name was chosen because of a formal resemblance between the two quantities, see Supplementary Information of [7].

Definition 10. *Entanglement* over partition $\mathscr{P} = \{M_1 \ldots M_m\}$ of $src(\mathfrak{m_D})$ is

$$\gamma(\mathfrak{m_D}, \mathscr{P}, d_{out}) = H\left[\mathfrak{m_D^\natural} \circ d_{out} \middle\| \bigotimes_{i=1}^{m} \pi_j \circ \mathfrak{m_j^\natural} \circ d_{out}\right]$$

where $\pi_j : \mathscr{V}S^{\mathbf{D}} \to \mathscr{V}S^{M_j}$ and $\mathfrak{m}_j = \{(k,l) \in \mathfrak{m_D} | k \in M_j\}$.

Projecting via π_j marginalizes onto the subspace $\mathscr{V}S^{M_j}$. Entanglement thus compares the measurement performed by the entire system with submeasurements over the decomposition of the source occasions into partition $\mathscr{P}$.

Theorem 9 (effective information decomposes additively when entanglement is zero)**.**

$$\gamma(\mathfrak{m_D}, \mathscr{P}, d_{out}) = 0 \implies ei(\mathfrak{m_D}, d_{out}) = \sum_{i=1}^{m} ei(\mathfrak{m}_j, d_{out}).$$

Proof: Follows from the observations that (i) $H[p\|p_1 \otimes p_2] = 0$ if and only if $p = p_1 \otimes p_2$; (ii) $H[p_1 \otimes p_2\|q_1 \otimes q_2] = H[p_1\|q_1] + H[p_2\|q_2]$; and (iii) the uniform distribution on $\mathbf{D}$ is a tensor of uniform distributions on subsystems of $\mathbf{D}$. ∎

The theorem shows the relationship between effective information and entanglement. If a system generates more information "than it should" (meaning, more than the sum of its subsystems), then the measurements it generates are entangled. Alternatively, only indecomposable measurements can be more precise than the sum of their submeasurements.

We conclude with some detailed computations for $X \times Y \overset{g}{\to} Z$, Diagram (11). Let $\mathscr{P} = \{X|Y\}$.

Theorem 10 (entanglement and effective information for $g : X \times Y \to Z$)**.**

$$\gamma(\mathfrak{m}_{XY}, \mathscr{P}, \delta_z) = \sum_{(x,y) \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \log_2 \frac{|g^{-1}(z)|}{|g^{-1}_{x \times Y}(z)| \cdot |g^{-1}_{X \times Y}(z)|}$$

$$= ei(\mathfrak{m}_{XY}, \delta_z) - ei(\mathfrak{m}_{X\bullet}, \delta_z) - ei(\mathfrak{m}_{\bullet Y}, \delta_z).$$

Proof: The first equality follows from Propositions 5 and 6

$$\gamma(\mathfrak{m}_{XY}, \mathscr{P}, \delta_z) = \sum_{(x,y) \in g^{-1}(z)} = \sum_{(x,y) \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \log_2 \left[\frac{1}{|g^{-1}(z)|} \cdot \frac{|g^{-1}(z)|}{|g^{-1}_{x \times Y}(z)|} \frac{|g^{-1}(z)|}{|g^{-1}_{X \times Y}(z)|}\right].$$

From the same propositions it follows that $ei(\mathfrak{m}_{XY}, \delta_z) - ei(\mathfrak{m}_{X\bullet}, \delta_z) - ei(\mathfrak{m}_{\bullet Y}, \delta_z)$ equals

$$\log_2 \frac{|X| \cdot |Y|}{|g^{-1}(x)|} - \sum_x \frac{|g^{-1}_{x \times Y}(z)|}{|g^{-1}(z)|} \log_2 \frac{|X| \cdot |g^{-1}_{x \times Y}(z)|}{|g^{-1}(z)|} - \sum_y \frac{|g^{-1}_{X \times y}(z)|}{|g^{-1}(z)|} \log_2 \frac{|Y| \cdot |g^{-1}_{X \times y}(z)|}{|g^{-1}(z)|}$$

$$= \log_2 \frac{1}{g^{-1}(z)} - \sum_{(x,y) \in g^{-1}(z)} \frac{1}{|g^{-1}(z)|} \cdot \log_2 \frac{|g^{-1}_{X \times y}(z)|}{|g^{-1}(z)|} \cdot \frac{|g^{-1}_{x \times Y}(z)|}{|g^{-1}(z)|}.$$

Entanglement quantifies how far the size of the pre-image of $g^{-1}(z)$ deviates from the sizes of its $X \times y$ and $x \times Y$ slices as x and y are varied. ∎

By Corollary 8 entanglement also equals $ei(\mathfrak{m}_{X\bullet} \to \mathfrak{m}_{XY}, \delta_z) - ei(\mathfrak{m}_{\bullet Y}, \delta_z)$. In Diagram (11) entanglement is the vertical arrow minus both arrows at the bottom, or the difference between opposing diagonal arrows. Note that the diagonal arrows from left to right are constructed by adding edge $v_Y \to v_Z$ to the null system and the subsystem $\mathfrak{m}_{X\bullet} = \{v_X \to v_Z\}$ respectively. Entanglement is the difference between the information generated by the diagonal arrows. It quantifies the difference between the information $\{v_Y \to v_Z\}$ generates in two different contexts.

Corollary 11 (characterization of disentangled set-valued functions). *Function $X \times Y \xrightarrow{g} Z$ performs a disentangled measurement when outputting z iff*

$$g^{-1}(z) = g^{-1}_{x \times Y}(z) \times g^{-1}_{X \times y}(z)$$

for any x, y such that $g(x, y) = z$.

Proof: By Theorem 10 entanglement is zero iff

$$|g^{-1}(z)| = |g^{-1}_{x \times Y}(z)| \cdot |g^{-1}_{X \times y}(z)|$$

for any x, y such that $g(x, y) = z$. This implies the desired result since $g^{-1}(z) \hookrightarrow g^{-1}_{x \times Y}(z) \times g^{-1}_{X \times y}(z)$. ∎

Thus, the measurement generated by g is disentangled iff its pre-image $g^{-1}(z)$ satisfies a strong geometric "rectangularity" constraint: that the pre-image decomposes into the product of its $x \times Y$ and $X \times y$ slices for all pairs of slices intersecting $g^{-1}(z)$. The categorizations performed within a disentangled measuring device have nothing to do with each other, so that the device is best considered as two (or more) distinct devices that happen to have been grouped together for the purposes of performing a computation.

Example 4. An XOR-gate $g : X \times Y \to Z$ outputting 0 generates an entangled measurement. The pre-image is $g^{-1}(0) = \{00, 11\}$ so the XOR-gate generates 1 bit of information about occasions v_X and v_Y. However, the bit is *not localizable*. The measurement generates no information about occasion v_X taken singly, since its output could have been 0 or 1 with equal probability; and similarly for v_Y.

Finally, and unsurprisingly, a function is completely disentangled across all its measurements iff it is a product of two simpler functions:

Corollary 12 (completely disentangled functions are products). *If $X \times Y \xrightarrow{g} Z$ is surjective, then $\gamma(\mathfrak{m}_{XY}, \mathscr{P}, \delta_z) = 0$ for all $z \in Z$ iff g decomposes into $X \times Y \xrightarrow{g_1 \times g_2} Z_1 \times Z_2 = Z$ for $X \xrightarrow{g_1} Z_1$ and $Y \xrightarrow{g_2} Z_2$.*
Proof: The reverse implication is trivial. In the forward direction, note that $Z = \{g^{-1}(z) | z \in Z\}$ and, by Corollary 11, each pre-image has product structure $g^{-1}(z) = g^{-1}_{x \times Y}(Z) \times g^{-1}_{X \times Y}(z)$. Let $Z_1 = \{g^{-1}_{X \times y} | y \in Y \text{ and } z \in Z\}$ and similarly for Z_2. Define

$$g_1 : X \to Z_1 : x \mapsto \text{the unique element of form } g^{-1}_{X \times y}(z) \text{ containing it,}$$

and similarly for g_2. ∎

6   Discussion

This paper developed techniques for analyzing the internal structure of distributed measurements. We introduced entanglement, which quantifies the extent to which a measurement is indecomposable. Entanglement can be shown to quantify context-dependence. Moreover, positive entanglement is necessary

for a system to generate more information than the sum of its subsystems. Along the way, we constructed the quale, which geometrically represents the compositional structure of a distributed measurement. The information-theoretic approach developed here is dual, in a precise sense, to the algorithmic perspective on computation. Studying duals $\mathfrak{m}^\natural$ instead of mechanisms $\mathfrak{m}$ shifts the focus from *what* the algorithm does to *how* it does it: instead of analyzing rules we analyze functional dependencies.

The intuition driving the paper is that the structure presheaf $\mathscr{F}$ is an information-theoretic analogue of a tangent space. A particle moving in a manifold X defines a vector field – a section of the tangent space to X, which is a sheaf. The tangent vector at a point depends on the particle's location at "nearby time-points": it is computed by taking the limit of difference in positions at t and $t+h$ as $h \to 0$. Similarly, a system performing a measurement generates a quale, a section of the structure presheaf consisting of "nearby counterfactuals". The quale is computed by applying Bayes' rule to determine which inputs could have led to the output.[3] How far this analogy can be developed remains to be seen.

Entanglement can be loosely considered as an information-theoretic analogue of curvature: the extent to which interactions within a system "warp" sections of $\mathscr{F}$ away from a product structure. A related approach to geometrically analyzing the complexity of interactions was proposed in [4]. In fact, this project began as an attempt to reformulate [7] in terms of sheaf cohomology using ideas from [4]. We failed at the first step since the structure presheaf is not a sheaf. However, the failure was instructive since it is precisely the *obstruction* to forming a sheaf that is of interest since it is the obstruction (entanglement) that quantifies indecomposability and context-dependence, and only systems whose measurements are entangled are able to generate more information than the sum of their subsystems.

References

[1] Samson Abramsky & Bob Coecke (2009): *Categorical Quantum Mechanics*. In K Engesser, D M Gabbay & D Lehmann, editors: *Handbook of Quantum Logic and Quantum Structures: Quantum Logic*, Elsevier.

[2] S Amari & H Nagaoka (2000): *Methods of Information Geometry*. AMS Translations of Mathematical Monographs 191, Oxford University Press.

[3] DJ Amit (1989): *Modelling brain function: the world of attractor neural networks*. Cambridge University Press.

[4] N Ay, E Olbrich, N Bertschinger & J Jost (2006): *A unifying framework for complexity measures of finite systems*. In: *Proceedings of ECCS06, European Complex Systems Society*, Oxford, UK, pp. ECCS06–174.

[5] David Balduzzi (2011): *Detecting emergent processes in cellular automata with excess information*. preprint .

[6] David Balduzzi & Giulio Tononi (2008): *Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework*. PLoS Comput Biol 4(6), p. e1000091, doi:10.1371/journal.pcbi.1000091.

[7] David Balduzzi & Giulio Tononi (2009): *Qualia: the geometry of integrated information*. PLoS Comput Biol 5(8), p. e1000462, doi:10.1371/journal.pcbi.1000462.

[8] J S Bell (1990): *Against 'Measurement'*. Physics World August, pp. 33–40.

[9] ER Berlekamp, JC Conway & RK Guy (1982): *Winning Ways for your Mathematical Plays*. 2, Academic Press.

[10] P A M Dirac (1958): *The Principles of Quantum Mechanics*. Oxford University Press.

[11] Tobias Fritz (2009): *A presentation of the category of stochastic matrices*. arXiv:0902.2554v1 .

[12] Martin Gardner (1970): *Mathematical Games - The Fantastic Combinations of John Conway's New Solitaire Game, Life*. Scientific American 223, pp. 120–123.

---

[3]A counterfactual input is "nearby" to an output if it causes (leads to) that output.

[13] M Giry (1981): *A categorical approach to probability theory*. In B Banaschewski, editor: *Categorical Aspects of Topology and Analysis*, Springer.

[14] G 't Hooft (1999): *Quantum gravity as a dissipative deterministic system*. Classical and Quantum Gravity 16(10).

[15] JJ Hopfield (1982): *Neural networks and physical systems with emergent computational properties*. Proc. Nat. Acad. Sci. 79, pp. 2554–2558.

[16] E T Jaynes (1957): *Information theory and statistical mechanics*. Phys. Rev. 106(4), pp. 620–630.

[17] E T Jaynes (1985): *Entropy and Search Theory*. In CR Smith & WT Grandy, editors: *Maximum-entropy and Bayesian Methods in Inverse Problems*, Springer.

[18] Stuart Kauffman, Carsten Peterson, Björn Samuelsson & Carl Troein (2003): *Random Boolean network models and the yeast transcriptional network*. Proc Natl Acad Sci U S A 100(25), pp. 14796–9, doi:10.1073/pnas.2036429100.

[19] S MacLane & Ieke Moerdijk (1992): *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer.

[20] P Panangaden (1998): *Probabilistic relations*. In C Baier, M Huth, M Kwiatkowska & M Ryan, editors: *PROBMIV'98*, pp. 59–74.