Categories of empirical models

Martti Karvonen

School of Informatics
University of Edinburgh
`martti.karvonen@ed.ac.uk`

A notion of morphism that is suitable for the sheaf-theoretic approach to contextuality is developed, resulting in a resource theory for contextuality. The key features involve using an underlying relation rather than a function between measurement scenarios, and allowing for stochastic mappings of outcomes to outcomes. This formalizes an intuitive idea of using one empirical model to simulate another one with the help of pre-shared classical randomness. This allows one to reinterpret concepts and earlier results in terms of morphisms. Most notably: non-contextual models are precisely those allowing a morphism from the terminal object; contextual fraction is functorial; Graham-reductions induce morphisms, reinterpreting Vorob'evs theorem; contextual models cannot be cloned.

1 Introduction

In the Abramsky-Brandenburger approach to contextuality [5] questions of non-locality and contextuality are formulated in the language of sheaf theory. The objects defined and studied in that approach are called empirical models. In this work, we present two notions of a morphism between empirical models, allowing us to further advance the use of categorical language in the study of contextuality. Our notion of a morphism can be seen as a formalization of the intuitive idea that we can use one empirical model to simulate another one - either deterministically or more generally using classical, pre-shared randomness to postprocess the measurement results. The key techniques are (i) letting the results of a measurement depend on a *set* of other measurements, (ii) incorporating stochastic postprocessing, using the Kleisli category of the distribution monad. These let us reinterpret various concepts and results from the literature in terms of the resulting category. For example, in Theorem 4.1 we characterize non-contextual models as those admitting a morphism from the terminal object.

If one thinks of empirical models as concrete experimental set-ups, one can give a concrete interpretation for the morphisms. A map $d \to e$ consists of a method to simulate the data that e would generate using only d. We will illustrate this idea with the following story: imagine a lab to perform $e \colon \langle X, \mathcal{M}, O \rangle$ where for each $x \in X$ there is a graduate student responsible for measuring x and a professor choosing which $C \in \mathcal{M}$ to execute on a given day. One night the students have a party at the lab, and all the equipment breaks down. Fearing the consequences, the students decide that experiments must continue without the professor noticing any abnormalities in the data generated. Luckily, the students happen to have friends at a lab containing an experimental set-up for d, and the students make a plan: when a measurement x is called for, the student responsible for it calls each student responsible for a measurement in some set $\pi(x)$ of measurements at the lab for d, and then outputs a result $o \in O$ based on the outcome. If the students manage to fool the professor (who knows what the data should look like) no matter which run $C_1, \ldots, C_n \in \mathcal{M}$ of experiments she chooses, the protocol defines a map $d \to e$, i.e. a way of simulating e using d.

However, one gets different notions of a map depending on the powers given to students. In particular, the fabricated outcome for x might depend on the measurement of $\pi(x)$ either deterministically or

stochastically, using classical and pre-shared randomness. There are also other powers student might be given that we do not study in this work. For example, the students might be allowed to randomize which measurements they are calling upon based on earlier measurement results, or they might be allowed to use a fixed quantum resource.

1.1   Related work

Notions of morphisms between empirical models or measurement scenarios have been defined before, e.g. in [23] and in [17]. The key features we add to obtain more general morphisms are (i) stochasticity and (ii) morphisms with an underlying simplicial relation rather than just a simplicial function. This additional generality is needed for the main results.

Somewhat similar ideas are developed in [8], and while not written with the sheaf-theoretic language, a rough comparison is possible: given a box (more or less corresponding to an empirical model), they consider the set of boxes one can obtain by pre- and postprocessing with non-contextual boxes. On one hand, their framework allows for probabilistic preprocessing, which we leave for further work. On the other hand, the preprocessing step allows only for a single measurement to be called upon, whereas we allow the simulation of measurement to depend on multiple measurements.

2   Background

In this section, we set up some notation and briefly cover the basics of the sheaf-theoretic approach to contextuality. We suggest reading [5] for a more detailed introduction. The only major differences to usual presentations are that we work over a semifield instead of a semiring (to allow for conditional distributions) and that we recall the definition of the Kleisli category of a monad.

Definition 2.1. A (finite) measurement scenario is a tuple $\langle X, \mathcal{M}, O \rangle$, where X and O are finite sets of measurements and outcomes, respectively, and $\mathcal{M}$ is a *measurement cover*, i.e. an antichain $\mathcal{M} \subset \mathcal{P}(X)$ that covers X. The members of $\mathcal{M}$ are called (maximal) measurement contexts.

The intuition is that X encodes the set of possible measurements on some system. However, it might not be feasible to simultaneously measure any subset $A \subset X$. This is captured by the measurement cover $\mathcal{M}$ - a subset $A \subset X$ is jointly measurable precisely when it is contained in some maximal measurement context, i.e. there is $C \in \mathcal{M}$ such that $A \subset C$. In other words, the jointly measurable subsets form an abstract simplicial complex.

We routinely think of the pair $\langle X, \mathcal{M} \rangle$ as a simplicial complex rather than making a distinction between the measurement cover and the simplicial complex generated by it - after all, a simplicial complex can be given by listing all its faces or only its (inclusion) maximal faces. This geometric viewpoint turns out to be quite handy, so we will recall some terminology needed.

Definition 2.2. Given two simplicial complexes (given by their antichains of maximal faces) $\langle X, \mathcal{M} \rangle$ and $\langle Y, \mathcal{N} \rangle$, a relation $\pi \colon X \to Y$ is a *simplicial relation* if it maps faces to faces, i.e. , if for every $C \in \mathcal{M}$ we have $\pi C \subset D$ for some $D \in \mathcal{N}$.

Next, we recall some terminology that uses the outcome set. The mapping $U \mapsto O^U$ induces a functor (in fact a sheaf) $\mathcal{P}(X)^{\mathrm{op}} \to \mathbf{Set}$, that we call the sheaf of events. The action of the functor on an inclusion $U \subset V$ is given by restriction: $(s \colon V \to O) \mapsto (s|_U \colon U \to O)$. Elements of O^U are called *U-sections* and when $U = X$, they are called *global sections*. To keep the empirical picture in mind, one should think of a U-section as an assignment that gives for each measurement in U its outcome.

In the sheaf-theoretic framework, an empirical model is formalised as a family of distributions, each distribution being a distribution over the events in a measurement context. However, one wants some leeway in defining what a 'distribution' means, and hence we define them rather generally. Breaking from tradition, we are slightly less general than usual - we define distributions over semifields rather than semirings, because we want to be able to define conditional distributions. This does not exclude examples of interest that we are aware of, and only Theorem 4.7 uses this restriction. Recall that a semifield is a commutative semiring $(R, +, \cdot, 0, 1)$ in which non-zero elements have multiplicative inverses. Informally, a semifield is like a field except that substraction need not be possible, as in the semifield of nonnegative real numbers.

Definition 2.3. Let $(R, +, \cdot, 0, 1)$ be a semifield. A (finite) R-distribution on a set X is a function $e \colon X \to R$ such that

- $\operatorname{supp} e := \{x \in X \mid e(x) \neq 0\}$ is finite
- $\sum_{x \in X} e(x) = 1$

For set X, the set of R-distributions on it is denoted by $D_R(X)$. The independent product of R-distributions $e \in D_R(X)$, and $d \in D_R(Y)$ is the distribution $e \otimes d \in D_R(X \times Y)$ defined by by $(e \otimes d)(x, y) = e(x)d(y)$. When $e \colon X \times Y \to R$ is an R-distribution, we define $\operatorname{supp}_Y e := \{y \in Y \mid \sum_{x \in X} e(x, y) \neq 0\}$. For $y \in \operatorname{supp}_Y e$ we define the *conditional* R-distribution $e(-|y)$ by

$$e(x|y) := e(-|y)(x) = e(x, y) / \sum_{x \in X} e(x, y)$$

Taking $e(y)$ to be a shorthand for $\sum_{x \in X} e(x, y)$, the usual laws concerning conditional distributions hold, namely

$$e(x|y) = e(x, y) / e(y) \qquad \text{and} \qquad e(x, y) = e(x|y)e(y).$$

One can think of an R-distribution on x as a formal linear combination $\sum_i^n r_i x_i$ where the coefficients $r_i \in R$ sum to 1. The assignment $X \mapsto D_R(X)$ extends to a functor $D_R \colon \mathbf{Set} \to \mathbf{Set}$ by sending a function $f \colon X \to Y$ to the function $D_R(f) \colon D_R(X) \to D_R(Y)$ defined by

$$D_R(f)(e)(y) = \sum_{x \in f^{-1} y} e(x)$$

written in terms of formal convex combinations, we can equivalently write the action of D_R on functions as

$$D_R(f) \sum_i r_i x_i := \sum_i r_i f(x_i)$$

The functor D_R is in fact a monad (see [15] and references therein) on **Set**, with unit $\eta \colon \operatorname{id}_{\mathbf{Set}} \to D_R$ defined by

$$\eta_X(x)(y) = 1x = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases}$$

and multiplication $\mu \colon D_R \circ D_R \to D_R$ defined by "matrix multiplication", i.e.

$$\mu(\sum_i^n r_i e_i)(x) = \sum_i^n (r_i e_i(x))$$

We recall (see e.g. [11, Proposition 4.1.6] for more) the definition of a Kleisli category for a monad.

Definition 2.4. Let (T, μ, η) be a monad on a category $\mathbf{C}$. The *Kleisli-category of* T, denoted by $\mathbf{C_T}$ has the same objects as $\mathbf{C}$, but a morphism $A \to B$ in $\mathbf{C_T}$ is given by a morphism $A \to TB$ in $\mathbf{C}$. The composite of $f: A \to TB$ and of $g: B \to TC$ is given by $\mu \circ T(g) \circ f: A \to TB \to TTC \to TC$. The identity morphism $A \to A$ in $\mathbf{C_T}$ is given by $\eta_A: A \to TA$ in $\mathbf{C}$. There is a canonical functor $\mathbf{C} \to \mathbf{C_T}$ given by $f \mapsto \eta \circ f$.

For a semifield R, we write $\mathbf{Set_R}$ instead of $\mathbf{Set_{D_R}}$ to avoid double subscripts, and we write $F_R: Set \to Set_R$ for the canonical functor. We record for future use the fact that taking product distributions behaves well with the functor D_R.

Lemma 2.5. *The map* $(e, d) \mapsto e \otimes d$ *defines a natural transformation* $D_R(-) \times D_R(-) \to D_R(- \times -)$.

Proof. Given R-distributions $e = \sum_i r_i x_i$ and $d = \sum_j s_j y_j$, we calculate as follows:

$$D_R(f \times g)(e \otimes d) = D_R(f \times g) \sum_{i,j} r_i s_j (x_i, y_j)$$

$$= \sum_{i,j} r_i s_j (f x_i, g y_j) = \sum_i r_i f x_i \otimes \sum_j g y_j = (D_R(f)e) \otimes (D_R(g)d) \qquad \square$$

One routinely composes the functor $D_R: \mathbf{Set} \to \mathbf{Set}$ with the event sheaf $\mathscr{P}(X)^{\mathrm{op}} \to \mathbf{Set}$ to obtain a presheaf $D_R(O^-): \mathscr{P}(X)^{\mathrm{op}} \to \mathbf{Set}$, which assigns to each set U of measurements the set of R-distributions over U-sections. Given an inclusion $U \subset V$, we obtain a map $D_R(O^V) \to D_R(O^U)$, given by marginalization: $e \in D_R(O^V)$ is mapped to $e|_U \in D_R(O^U)$ defined by

$$e|_U s = \sum_{t \in O^V, t|_u = s} e(t).$$

Definition 2.6. Given a semifield R, an *R-empirical model over the measurement scenario* $\langle X, \mathscr{M}, O \rangle$ is given by *a compatible family* $e = (e_C)_{C \in \mathscr{M}}$ of R-distributions over O^C, meaning that $e_C|_{C \cap D} = e_D|_{C \cap D}$ for all $C, D \in \mathscr{M}$, i.e. the distributions e_C agree on overlaps. If $U \subseteq C$ for some $C \in \mathscr{M}$, we write $e|_U$ instead of $e_C|_U$: after all, $e|_U$ does not depend on the choice of $C \in \mathscr{M}$. We denote e being an empirical model over the measurement scenario $\langle X, \mathscr{M}, O \rangle$ by $e : \langle X, \mathscr{M}, O \rangle$.

An *R-empirical model* $e : \langle X, \mathscr{M}, O \rangle$ is

- *R-non-contextual* if there is $d \in D_R(O^X)$ such that $e_C = d|_C$ for all $C \in \mathscr{M}$, i.e. if there exists a distribution on global sections explaining it.

- *R-contextual* if it is not R-non-contextual.

- *strongly R-contextual* if there is no global section consistent with it, i.e. if there is no $s \in O^X$ such that $s|_C \in \mathrm{supp}(e_C)$ for all $C \in \mathscr{M}$

Given R-empirical models $e^1, \ldots, e^n : \langle X, \mathscr{M}, O \rangle$ over the same measurement scenario and $r_1, \ldots, r_n \in R$ such that $\sum_{i=1}^n r_i = 1$, we define the R-empirical model $\sum_{i=1}^n r_i \cdot e^i$ by $(\sum_{i=1}^n r_i \cdot e^i)_C = \sum_{i=1}^n r_i \cdot e_C^i$. If each e^i is no-signalling, then so is $\sum_{i=1}^n r_i \cdot e^i$.

The *non-contextual fraction* of an $\mathbb{R}^+$-empirical model e is defined as the maximal $\lambda \in [0, 1]$ such that e decomposes as $e = \lambda e^{NC} + (1 - \lambda)e'$ where e^{NC} and e' are empirical models with e^{NC} non-contextual. It is denoted by $NCF(e)$, and the *contextual fraction* of e is defined as $CF(e) := 1 - NCF(e)$.

Strictly speaking, the above definition defines so-called no-signalling empirical models, and general empirical models are obtained by dropping the compatibility conditions. Since we only work with no-signalling empirical models, we've chosen to just call them empirical models.

In practice, R is usually one of the following semifields:

- The semifield $(\mathbb{R}^+, +, \cdot, 0, 1)$ of nonnegative reals, giving rise to finitely supported probability distributions. Usually $\mathbb{R}^+$-contextuality is either called probabilistic contextuality or just contextuality.

- The field $(\mathbb{R}, +, \cdot, 0, 1)$ of real numbers, giving rise to finitely supported *signed* measures. In [5] it is shown that being $\mathbb{R}$-non-contextual is equivalent to being no-signalling.

- The semifield $(\mathbb{B}, \vee, \wedge, 0, 1)$ of Booleans. Now finite $\mathbb{B}$-distributions on X correspond to finite subsets of X, and $\mathbb{B}$-contextuality is known as logical contextuality or possibilistic contextuality.

Since all non-signalling models over $\mathbb{R}$ are non-contextual and hence not strongly contextual, the notion of strong $\mathbb{R}$-contextuality is somewhat uninteresting. Hence one usually cares about strong contextuality over the positive reals or the booleans, and these notions are in fact equivalent, i.e. an $\mathbb{R}^+$-empirical model is strongly contextual iff its possibilistic collapse (defined later) is. Thus in the sequel we only talk about strong contextuality omitting the semifield R.

3   Defining morphisms

Before we give the full formal definitions, let us sketch them using the intuitive story from the introduction as guidance. First of all, any measurement $x \in X$ is assumed to be simulated by doing a set of measurements $\pi(x) \subset Y$. Because the students are supposed to work independently once the professor sets them to work, $\pi(x)$ should only depend on x and not on the set of other measurements performed with it. Hence we get a relation $\pi \colon X \to Y$. For the simulation to be successful, for any $C \in \mathcal{M}$ chosen, one must be able to perform the measurement $\pi(C)$, meaning that $\pi(C)$ has to be a jointly measurable subset of Y. In other words, π must be a simplicial relation.

Once π has been fixed, how about the rest of the simulation protocol? For any joint outcome $s \in P^{\pi C}$ for the measurements in $\pi(C)$, one must know what outcome in O^C to output. Because different measurements in X are measured by different students, possibly at different locations, we assume that the students cannot coordinate after each student responsible for measuring x receives the outcomes for measurements in $\pi(x)$. Any coordination (including any possible shared & classical randomness) must be done before various students are instructed to perform their measurements by the professor. Even worse, each student is just told to perform the measurement x without being revealed the full context containing x.

Thus the protocol in fact defines, for any $U \subset X$, a map $P^{\pi U} \to O^U$ (and if randomness is allowed, this map lives in the Kleisli category $\mathbf{Set_R}$). After all, if an arbitrary, not necessarily possible U was chosen and for each $x \in U$ the result for $\pi(x)$ was given, the students would still know which result O^U to output. Moreover, the students acting independently implies that the family of maps $P^{\pi U} \to O^U$ has to be a natural transformation: it should not matter whether the professor commissions the experiment V and then throws away some of the results to get a result $s \in O^U$, or whether the students are told to perform U (thus ignoring any results in $\pi(V) \setminus \pi(U)$) straight away.

Having been suitably motivated, we proceed to the actual definitions and discuss examples.

Definition 3.1. Let $\langle X, \mathcal{M}, O \rangle$ and $\langle Y, \mathcal{N}, P \rangle$ be measurement scenarios. A *deterministic morphism* $\langle Y, \mathcal{N}, P \rangle \to \langle X, \mathcal{M}, O \rangle$ consists of:

- a simplicial relation $\pi \colon X \to Y$;

- a natural transformation $\sigma \colon P^{\pi(-)} \to O^{(-)}$.

Given an empirical model $d\colon \langle Y, \mathcal{N}, P \rangle$, its pushforward along a deterministic morphism (σ, π) is the empirical model $\sigma_* d\colon \langle X, \mathcal{M}, O \rangle$ defined by

$$\sigma_* d_C = D_R(\sigma)(d|_{\pi(C)})$$

Let $e\colon \langle X, \mathcal{M}, O \rangle$ and $d\colon \langle Y, \mathcal{N}, P \rangle$ be empirical models. Then a *deterministic simulation* $d \to e$ consists of a deterministic morphism $\langle Y, \mathcal{N}, P \rangle \to \langle X, \mathcal{M}, O \rangle$ such that

$$e = \sigma_* d$$

The category of R-empirical models and deterministic simulations is denoted by **EmpDet$_R$**.

This definition can be motivated from the mathematical point of view as well, for example, via an analogy to algebraic geometry[1]. Roughly speaking, a scheme is a pair $(X, \mathcal{O}_X)$, where X is a topological space and $\mathcal{O}_X$ is a sheaf of local rings on it, and a morphism of schemes $(X, \mathcal{O}_X) \to (Y, \mathcal{O}_Y)$ consists of a continuous map $\pi\colon X \to Y$ and of a natural transformation $\mathcal{O}_Y \to \pi_* \mathcal{O}_X$ subject to some axioms. Similarly, a measurement scenario comes with an associated simplicial complex and with a presheaf of outcomes/distributions on the simplices, so it is reasonable to expect that a morphism of measurement scenarios consists of an underlying map between simplicial complexes and of a transformation between the two presheaves (suitably composed with the map of simplicial complexes). Moreover, an empirical model is a compatible family over the presheaf. Hence a morphism of empirical models should be a morphism between the underlying measurement scenarios taking one compatible family to the other.

Remark 3.2. We did not restrict ourselves to no-signalling models merely because of disinterest: the definition above would not work as-is without no-signalling: after all, for the pushforward to be well-defined, $d|_{\pi(C)}$ should not depend on which maximal context containing $\pi(C)$ is chosen, as long as there is one.

Lemma 3.3. *[Deterministic morphisms can be glued together] Let $\langle X, \mathcal{M}, O \rangle$ and $\langle Y, \mathcal{N}, P \rangle$ be measurement scenarios and $\pi\colon X \to Y$ a simplicial relation. Let $U_1, \ldots U_n$ be a cover of X and assume we are given a compatible family of natural transformations $\sigma^i\colon P^{\pi(-)} \to O^-$, i.e. each σ^i is a map of presheaves on U_i, and the σ^i agree on overlaps, meaning that $\sigma^i_{U_i \cap U_j} = \sigma^j_{U_i \cap U_j}$ for every i, j. Then there is a unique morphism (π, σ) of measurement scenarios such that σ restricts to σ^i on each $\mathscr{P}(U_i)$.*

Proof. The presheaf O^- is in fact a sheaf, so this is a consequence of the fact that morphisms of sheaves glue. See e.g. [16, Proposition 2.8.1.] $\qquad\square$

Example 3.4. Consider the $\mathbb{R}^+$-empirical model $e\colon \langle \{x,y\}, \{\{x,y\}\}, \{0,1\} \rangle$ given by $e(x \mapsto 1, y \mapsto 0) = e(x \mapsto 0, y \mapsto 1) = 1$ and $e(s) = 0$ otherwise, i.e. e consists of two perfectly anticorrelated flips of a fair coin. Then all the information about the second coin flip is already (deterministically) present in the first, and this is made precise by the deterministic simulation $(\pi, \sigma)\colon e|_x \to e$ given by $\pi(x) = \pi(y) = x$ and $\sigma_{\{x,y\}}(x \mapsto a) = (x \mapsto a, y \mapsto (1-a))$.

Example 3.5 (Restriction)**.** If $e\colon \langle X, \mathcal{M}, O \rangle$ is an empirical model and $i\colon Y \hookrightarrow X$ is a subset of X with inclusion $i\colon Y \to X$, one gets a restricted empirical model $e|_Y\colon \langle Y, \{C \cap Y | C \in \mathcal{M}\}, O \rangle$ defined by $(e|_Y)_{C \cap Y} = e_{C \cap Y}$. One can obviously simulate the restricted empirical model using the larger one, and this corresponds to the morphism $(i, \mathrm{id})\colon e \to e|_Y$.

---

[1]This analogy is not perfect, of course. For one, we define our morphisms to go in the opposite direction compared to the underlying map between simplicial complexes. This is to guarantee that "points" of e, i.e. maps from the terminal object to e correspond to distributions on global sections explaining e, but one could have reasonably chosen the opposite convention. Perhaps a more surprising difference is that instead of the direct image sheaf we use the inverse image sheaf. Since in our set-up all topological spaces are finite and discrete, using the (in general more complicated) inverse image functor poses no problems.

Example 3.6 (Coarse-graining). If $f\colon O \to P$ is surjective, one can define, for $e\colon \langle X, \mathcal{M}, O \rangle$ the coarse-graining of e along f as the empirical model $e/f\colon \langle X, \mathcal{M}, P \rangle$, defined via $e/f_C(s) := \sum_{t\colon C\to O, ft=s} e_C(t)$. Applying f to the outcome lets one simulate e/f using e, and this is made precise by the morphism $(\mathrm{id}, \sigma)\colon e \to e/f$, where σ is the natural transformation induced by the function f.

Example 3.7. We now exhibit a situation where π is not a function but a relation. Consider the measurement scenario $\langle \{x,y,z\}, \{\{x,y,z\}\}, \{0,1\} \rangle$ of three jointly measurable binary variables and let e be $\mathbb{R}^+$-empirical model over it corresponding to the variables x and y being independent flips of a fair coin and z being the total number of heads modulo 2. Then there is no deterministic simulation $(\pi, \sigma)\colon e|_{\{x,y\}} \to e$ where π is a function, because one cannot infer the value of z knowing the value of only one of x, y. Informally, one can clearly simulate e by $e|_{\{x,y\}}$: to know the value of z, measure both x and y and calculate, and the values of x and y are given by themselves. This is made formal by the deterministic simulation $(\pi, \sigma)\colon e|_{\{x,y\}} \to e$ where π is given by $\pi(z) = \{x,y\}$ and $\pi|_{\{x,y\}} = \mathrm{id}$ and $\sigma_{\{x,y\}} = \mathrm{id}$, $\sigma_z(x \mapsto a, y \mapsto b) = z \mapsto (a+b \mod 2)$ (note that by Lemma 3.3 this is enough to specify σ).

Example 3.8. In [18] the authors show that in any $(2,2,l)$ or $(2,k,2)$ Bell scenario, the occurrence of Hardy's paradox is a necessary and sufficient condition for the model to be possibilistically contextual. The crucial thing to note is that here "occurrence" means finding a certain partially filled table inside the empirical model in question. One can make sense of this in our framework by saying that there are several versions of Hardy's paradox - one for each of the ways of completing the table - and being possibilistically contextual is equivalent to being able to (deterministically) simulate one of these.

Definition 3.9. Let $\langle X, \mathcal{M}, O \rangle$ and $\langle Y, \mathcal{N}, P \rangle$ be measurement scenarios. A *(R-stochastic) morphism* $\langle Y, \mathcal{N}, P \rangle \to \langle X, \mathcal{M}, O \rangle$ consists of

- a simplicial relation $\pi\colon X \to Y$;
- a natural transformation $\sigma\colon P^{\pi(-)} \to D_R(O^{(-)})$.

The composite of $(\pi, \sigma)\colon \langle Z, \mathcal{K}, Q \rangle \to \langle Y, \mathcal{N}, P \rangle$ and $(\rho, \tau)\colon \langle Y, \mathcal{N}, P \rangle \to \langle X, \mathcal{M}, O \rangle$ is given by $(\rho \circ \pi, \mu \circ D_R(\rho) \circ \tau)$ i.e. by ordinary composition in the first variable and Kleisli composition in the second. Given an empirical model $d\colon \langle Y, \mathcal{N}, P \rangle$, its pushforward along a morphism (σ, π) is the empirical model $\sigma_* d\colon \langle X, \mathcal{M}, O \rangle$ defined by

$$\sigma_* d_C = \mu D_R(\sigma)(d|_{\pi(C)})$$

Let $e\colon \langle X, \mathcal{M}, O \rangle$ and $d\colon \langle Y, \mathcal{N}, P \rangle$ be empirical models. Then a *simulation* $d \to e$ consists of an R-stochastic morphism $\langle Y, \mathcal{N}, P \rangle \to \langle X, \mathcal{M}, O \rangle$ such that

$$e = \sigma_* d$$

Identity on d is given by (id, η) and the composite of $(\pi, \sigma)\colon d \to e$ and $(\rho, \tau)\colon e \to f$ is given by $(\rho \circ \pi, \mu \circ D_R(\rho) \circ \tau)$ i.e. by ordinary composition in the first variable and Kleisli composition in the seconds. The category of R-empirical models and simulations is denoted by **Emp$_R$**.

Remark 3.10. In the previous definition, instead of a natural transformation $\sigma\colon P^{\pi(-)} \to D_R(O^{(-)})$ we could have equivalently asked for a natural transformation $\sigma\colon F_R(P^{\pi(-)}) \to F_R(O^{(-)})$, i.e. for a natural transformation between two functors $\mathscr{P}(X)^{\mathrm{op}} \to \mathbf{Set_R}$. This makes it apparent that composition is associative.

Given the terminology, one would expect deterministic morphisms (deterministic simulations) to be special cases of morphisms (simulations). Indeed, this is the case - one just needs to apply the inclusion $F_R\colon \mathbf{Set} \to \mathbf{Set_R}$ to the Kleisli category. To be more specific, if (π, σ) is a deterministic morphism, then $(\pi, F_R\sigma)$ is a morphism. That the same claim holds for simulations boils down to the fact that the pushforward of an empirical model does not depend on whether one uses (π, σ) or $(\pi, F_R\sigma)$. We note that for a simulation $(\pi, \sigma)\colon d \to e$ one only needs the parts of d that π maps to.

Lemma 3.11. *Any simulation* $(\pi, \sigma)\colon d \to e$ *factors through the restriction* $d \to d|_{\pi(X)}$.

Unfortunately Lemma 3.3 fails for R-stochastic simulations. This is precisely because $D_R(O^-)$ fails to be a sheaf. However, one can glue together maps (though not uniquely) along partitions. It is straight-forward to extend the following lemma to an n-ary partition $(U_1, \ldots U_n)$.

Lemma 3.12. *Let* $\langle X, \mathcal{M}, O \rangle$ *and* $\langle Y, \mathcal{N}, P \rangle$ *be measurement scenarios and* $\pi\colon X \to Y$ *a simplicial relation. If* (U_1, U_2) *is a partition of* X *and* $\sigma^i\colon P^{\pi(-)} \to D_R(O^-)$ *is a natural transformation between presheaves on* U_i *for* $i = 1, 2$, *then there is a natural transformation* $\sigma^1 \otimes \sigma^2\colon P^{\pi(-)} \to D_R(O^-)$ *that restricts to* σ_i *on* $\mathscr{P}(U_i)$, *defined by*

$$(s\colon \pi(V) \to P) \mapsto (\sigma^1 \otimes \sigma^2)_V(s) := \sigma^1_{V \cap U_1}(s|_{\pi(V \cap U_1)}) \otimes \sigma^2_{V \cap U_2}(s|_{\pi(V \cap U_2)})$$

Proof. The morphism $\sigma^1 \otimes \sigma^2$ factors as a composite of natural transformations

$$P^{\pi(-)} \to P^{\pi(- \cap U_1)} \times P^{\pi(- \cap U_2)} \to D_R(O^{- \cap U_1}) \times D_R(O^{- \cap U_2}) \to D_R(O^{- \cap U_1} \times O^{- \cap U_2}) \cong D_R(O^-)$$

where the first natural transformation is the pairing of the restrictions $P^{\pi(-)} \to P^{\pi(- \cap U_i)}$, the second is the cartesian product of the natural transformations σ^i, the third is the the product distribution map that is natural by Lemma 2.5, and the final isomorphism stems from the fact that $O^{- \cap U_1} \times O^{- \cap U_2} \cong O^-$. $\quad\square$

Proposition 3.13. *Let* $(\pi, \sigma^i)\colon d \to e_i$ *be simulations, where each* e_i *is an empirical model over the same measurement scenario. Then* $(\pi, \sum_{i=1}^n r_i \sigma^i)$ *defines a map* $d \to \sum_{i=1}^n r_i e_i$, *where* $\sum_{i=1}^n r_i \sigma^i$ *is defined by*

$$\left(\sum_{i=1}^n r_i \sigma^i\right)_U(s) := \sum_{i=1}^n r_i \sigma^i_U(s)$$

Proof. It is not hard to show that convex combinations are preserved by the restriction maps $D_R(O^V) \to D_R(O^U)$ induced by inclusions $U \to V$, i.e. that $(\sum r_i e_i)|_U = \sum r_i(e_i|_V)$. Since each σ_i is natural, this implies that $\sum r_i \sigma_i$ is as well. Hence it remains to see that the pusforward of d along $\sum r_i \sigma_i$ is $\sum r_i e_i$. But this is not hard either:

$$\left(\sum r_i \sigma^i\right)_* d = \sum r_i(\sigma^{i*} d) = \sum r_i e_i \qquad\qquad\qquad\square$$

Example 3.14. This time we discuss the examples more informally, leaving the formal description of the morphisms as an exercise. Consider again two jointly measurable variables. The first one is a flip of a fair coin, whereas the second one depends on the first as follows: if the first coin is heads, the second coin flip is fair, whereas if the first coin is tails, the second coin is flipped with a bias of $2/3$. Clearly knowing just the value of the first coin flip lets you simulate the experiment stochastically - just flip a coin with the appropriate bias. This defines a simulation $e|_x \to e$. One can go even further and simulate e stochastically without using e at all —just perform the whole experiment yourself. This ability to simulate e from thin air corresponds to non-contextuality of e, as will be seen in Theorem 4.1.

4   Reinterpreting contextuality with morphisms

The category $\mathbf{Emp_R}$ has a terminal object 1 given by the unique R-empirical model on the empty measurement scenario $\langle \emptyset, \emptyset, O \rangle$.[2] Admitting a morphism from the terminal object is equivalent to non-contextuality.

---

[2]It is easy to check that the choice of the outcome set does not matter.

Theorem 4.1. *Distributions on global sections explaining e are in one-to-one correspondence with simulations $1 \to e$. In particular, an R-empirical model is non-contextual iff there is a simulation $1 \to e$.*

Proof. Given an R-empirical model $e \colon \langle X, \mathcal{M}, O \rangle$, the data for a simulation $1 \to e$ consists of a simplicial relation $X \to \emptyset$, of which there is only one, and of a natural transformation from the constant singleton presheaf to $D_R(O^-)$. Hence, the data for a simulation $1 \to e$ corresponds to a probability distribution in $D_R(O^X)$. This data is in fact a simulation iff it restricts to e_C for each $C \in \mathcal{M}$. $\qquad \square$

Let $f \colon R \to S$ be a homomorphism of semifields. Then f induces a natural transformation (in fact, a monad homomorphism) $\kappa(f) \colon D_R \to D_S$ defined by

$$ e \in D_R(X) \mapsto \kappa(f)_X(e) \in D_S(X) \qquad\qquad \kappa(f)_X(e)(s) := f(e(s)) $$

This is known to map R-empirical models to S-empirical models by sending the R-empirical model $e \colon \langle X, \mathcal{M}, O \rangle$ to the S-empirical model $F_f(e) \colon \langle X, \mathcal{M}, O \rangle$ defined by $F_f(e)_C = \kappa(f)(e_C)$.

Theorem 4.2. *The map $e \mapsto F_f(e)$ extends to a functor $\mathbf{Emp_R} \to \mathbf{Emp_S}$. Hence S-contextuality of $F_f(e)$ implies R-contextuality of e. In particular, logical contextuality implies probabilistic contextuality.*

Proof. The action of F_f on an R-stochastic simulation (π, σ) is defined by

$$ F_f(\pi, \sigma) := (\pi, \kappa(f) \circ \sigma) $$

It is straightforward to check that this defines a functor $\mathbf{Emp_R} \to \mathbf{Emp_S}$. By construction it preserves the terminal object, so that if there is a map $1 \to e$ in $\mathbf{Emp_R}$, then there must be one in $\mathbf{Emp_S}$. $\qquad \square$

The possibilistic collapse of an $\mathbb{R}^+$-empirical model is defined to be the image of e under the functor $F \colon \mathbf{Emp_{\mathbb{R}^+}} \to \mathbf{Emp_{\mathbb{B}}}$ induced by the unique semiring homomorphism $\mathbb{R}^+ \to \mathbb{B}$. It is known not to be surjective on objects [1, Proposition 9.1.] and neither is it full - for example, there are probabilistically contextual models that are not logically contextual, so that there is a map $1 \to F(e)$ in $\mathbf{Emp_{\mathbb{B}}}$ not arising as the image of a map $1 \to e$ in $\mathbf{Emp_{\mathbb{R}^+}}$

While R-contextuality admits a categorical description, we have not been able to obtain a similar one for strong contextuality. However, morphisms do respect strong contextuality.

Theorem 4.3. *Let $(\pi, \sigma) \colon d \to e$ be an R-stochastic simulation where R is either $\mathbb{R}^+$ or $\mathbb{B}$. If e is strongly contextual, d is as well.*

Proof. We prove the contrapositive, so assume that $d \colon \langle Y, \mathcal{N}, P \rangle$ is not strongly contextual. Then there exists a global section $s \in P^Y$ such that $s|_D \in \mathrm{supp}(d_D)$ for all $D \in \mathcal{N}$. Then $e = \sigma_* d$ implies that any global section $p \in O^X$ in the support of $\sigma_X(s)$ satisfies $p|_C \in \mathrm{supp}\, e_C$ for every $C \in \mathcal{M}$, so that e is not strongly contextual. $\qquad \square$

Lemma 4.4. *Pushforward commutes with convex mixtures: if $d_1, \ldots d_n \colon \langle X, \mathcal{M}, O \rangle$ are empirical models, $r_1 \ldots r_n \in R$ sum to 1 and $(\pi, \sigma) \colon \langle X, \mathcal{M}, O \rangle \to \langle Y, \mathcal{N}, P \rangle$ is an R-stochastic simulation, then*

$$ \sigma_* \left(\sum r_i d_i \right) = \sum r_i \sigma_*(d_i) $$

Proof. This can be established via a straightforward calculation, or via abstract nonsense. We opt for the latter: objects in the Kleisli category are the free D_R-algebras and hence $\mu D_R(\sigma)$ is a homomorphism of D_R-algebras, which means that the equation to be proved holds. $\qquad \square$

Theorem 4.5. *Non-contextual fraction defines a functor* $\mathbf{Emp}_{\mathbb{R}^+} \to [0,1]$, *where* $[0,1]$ *is viewed as a category via its order relation.*

Proof. Let $(\pi,\sigma)\colon d \to e$ be a simulation, and consider any decomposition of d as $\lambda d^{NC} + (1-\lambda)d'$, with d^{NC} non-contextual. Then

$$e = \sigma_* d = \sigma_*(\lambda d^{NC} + (1-\lambda)d') = \lambda \sigma_* d^{NC} + (1-\lambda)\sigma_* d'$$

where the last equation uses Lemma 4.4. Now $\sigma_* d^{NC}$ is non-contextual and hence the convex decomposition of e obtained implies that $NCF(d) \le NCF(e)$. $\square$

Definition 4.6. Let $\langle X, \mathscr{M} \rangle$ be a simplicial complex. Given a vertex $x \in X$ that belongs to exactly one maximal face, we say that there is a *Graham-reduction* from $\langle X, \mathscr{M} \rangle$ to the subcomplex $\langle X \setminus \{x\}, \{C \setminus \{x\} \mid C \in \mathscr{M}\}\rangle$. If there is a Graham-collapse from $\langle X, \mathscr{M} \rangle$ to $\langle Y, \mathscr{N} \rangle$, we write $\langle X, \mathscr{M} \rangle \rightsquigarrow \langle Y, \mathscr{N} \rangle$. We call a simplicial complex $\langle X, \mathscr{M} \rangle$ *acyclic* if there is a sequence of Graham collapses from $\langle X, \mathscr{M} \rangle$ to the empty simplicial complex.

Vorob'evs theorem [22] is already known (see [6, p.13]) to imply that for questions about contextuality one can consider the reduced model. This can be recast and proved in terms of simulations.

Theorem 4.7. *Graham-reductions induce morphisms, i.e. if* $e\colon \langle X, \mathscr{M}, O \rangle$ *is an empirical model and* $\langle X, \mathscr{M} \rangle \rightsquigarrow \langle X \setminus \{x\}, \{C \setminus \{x\} \mid C \in \mathscr{M}\}\rangle$ *is a Graham-reduction, then there is a simulation* $e|_{X \setminus \{x\}} \to e$. *Hence any empirical model with an acyclic underlying simplicial complex is non-contextual.*

Proof. We define a simulation $e|_{X \setminus \{x\}} \to e$ in the appendix using Lemma 3.12. $\square$

The author has not been able to find monoidal structures on all of $\mathbf{Emp}_{\mathbf{R}}$. However, if one defines $\mathbf{Emp}_{\mathbf{R,O}}$ to be the full subcategory consisting of all empirical models taking outcomes in a fixed set O, there is one. This extends the operation

$$e\colon \langle X, \mathscr{M}, O \rangle, d\colon \langle Y, \mathscr{N}, O \rangle \mapsto e \otimes d \colon \langle X \sqcup Y, \mathscr{M} * \mathscr{N} := \{C \sqcup D \mid C \in \mathscr{M}, D \in \mathscr{N}\}, O \rangle$$

defined in [3] to morphisms. One should think of this operation as doing experimental set-ups and simulations independently in parallel. Formally, it is defined by

$$(e \otimes d)_{C \sqcup D} = e_C \otimes d_D \qquad (\pi,\sigma) \otimes (\rho,\tau) = (\pi \sqcup \rho, \sigma \otimes \tau),$$

where $\sigma \otimes \tau$ is the natural transformation built using Lemma 3.12. It is straightforward but tedious to check that this results in a symmetric monoidal structure on $\mathbf{Emp}_{\mathbf{R,O}}$ with the tensor unit given by the terminal object $1\colon \langle \emptyset, \emptyset, O \rangle$. One might view these SMCs as resource theories (in the sense of [12]) for contextuality. From this point of view Theorem 4.5 amounts to saying that the non-contextual fraction is a monotone for our resource theory. Next we will prove that this resource-theory is no-cloning, i.e. that no contextual empirical model can be cloned using only classical correlations.

Theorem 4.8 (No-cloning). *There is a simulation* $e \to e \otimes e$ *iff there is a simulation* $1 \to e$, *i.e. if* e *is non-contextual.*

Proof. The parallel composition of a morphism $1 \to e$ with itself gives rise to the composite

$$e \to 1 \cong 1 \otimes 1 \to e \otimes e,$$

proving the easier direction.

For the converse, consider an empirical model $e\colon \langle X, \mathscr{M}, O \rangle$ that can be cloned freely, i.e. admits a simulation $e \to e \otimes e$. We use induction on $|X|$ to build a simulation $1 \to e$, the base case $X = \emptyset$ being clear. Assuming the claim holds for all $m < |X| =: n$, we proceed. Note that from $e \to e \otimes e$ and monoidal operations we can build a simulation $(\pi, \sigma)\colon e \to e^{\otimes n}$. Now $\pi\colon \langle X^{\sqcup n}, \mathscr{M}^{*n} \rangle \to \langle X, \mathscr{M} \rangle$. Define $(\pi_i, \sigma^i)\colon e \to e$ as the composite $e \to e^{\otimes n} \to e^{\otimes n}|_{X_i} \cong e$, where the morphism in the middle is the restriction to ith copy of e in $e^{\otimes n}$. Now $\pi\colon X^{\sqcup n} \to X$ being simplicial means that $\bigcup_{i=1}^{n} \pi(C_i)$ is a jointly measurable subset whenever each C_i is. We split into two cases:

If $\pi_i X = X$ for each i, then, enumerating X as $x_1, \dots x_n$, we can choose y_i such that $x_i \in \pi_i(y_i)$ for each i. Hence $X = \cup \{x_i\}$ is jointly measurable, so that e is non-contextual.

Otherwise $\pi_i X \subsetneq X$ for some i. Then (π_i, σ^i) restricts to a morphism $e|_{\pi_i(X)} \to e$ using Lemma 3.11. Now, doing the restriction $e \to e|_{\pi_i(X)}$ in parallel results in the composite

$$e|_{\pi_i(X)} \to e \to e \otimes e \to e|_{\pi_i(X)} \otimes e|_{\pi_i(X)}.$$

Hence by the induction assumption $e|_{\pi_i(X)}$ is non-contextual, so take the composite $\emptyset \to e|_{\pi_i(X)} \to e$. $\qquad\square$

5   Further questions

Structure of $\mathbf{Emp_R}$ The main thread going through this work is that the existence of a simulation $d \to e$ implies that e is at most as contextual as d is. Assuming that our definition fully captures the intuitive idea of "simulation with classical randomness" (see more below), one could argue that the study of the contextuality hierarchy amounts to studying the category $\mathbf{Emp_R}$. This motivates cataloguing the categorical properties of $\mathbf{Emp_R}$. For instance, does the factorization from Lemma 3.11 give rise to a (co)pure-(co)mixed factorization system in the sense of [13]?

Likewise, Theorem 4.1 guarantees that non-contextual models can be defined in terms of the category $\mathbf{Emp_R}$. Can other interesting classes of empirical models be recognised in terms of the category $\mathbf{Emp_R}$? This question applies most pressingly to strongly contextual models, but can also be asked about models admitting an AvN-argument [2] or a cohomological obstruction [6], or perhaps more boldly, about models admitting a quantum realisation. A weaker variant is to ask if we can transport AvN-arguments or cohomological obstructions along morphisms. Similarly, in [3], several formulas concerning the contextual fraction are proved, and it would be interesting to see if one can explain those formulas as categorical properties of the functor CF.

Relationship with other approaches to contextuality.   There are several alternative approaches contextuality, for example, based on operational equivalence [19], hypergraphs [7] or effect-algebras [20]. It would be interesting to relate these formalisms to each other via categorical isomorphisms, equivalences or at least adjunctions. Some work in this direction already exists: for example, [23] defines a category of empirical models and a category suitable for the equivalence-based approach, and proves an isomorphism between the two. However, the morphisms on both sides are deterministic[3]. Similarly, in [7, Appendix D] a bijection between empirical models in our sense and probabilistic models on hypergraphs is defined. However, the question whether this extends to morphisms is not investigated. Moreover, the most obvious definition for morphisms in the combinatorial approach comes from morphisms of hypergraphs, and as such is deterministic. The same drawback applies to the effect-algebraic approach [20]:

---

[3]Stochastic morphisms with preprocessing but without dependence on multiple measurements in the equivalence-based setting are investigated in [14]. Since we do not have preprocessing but have dependencies on several measurements, hoping for an isomorphism might be too much.

empirical models are regarded as maps from an effect algebra E to the unit interval. In other words, the most natural category to work in is the slice category $\mathbf{EA}/[\mathbf{0}, \mathbf{1}]$, but then morphisms of empirical models seem to boil down to deterministically mapping effects to effects (i.e. outcomes to outcomes) in a way that preserves the probabilities. Of course, this is not meant as a dismissal of the other approaches: rather, it is a call to define morphisms in the appropriate generality for each of the approaches and to see how the resulting categories relate to each other.

Analogy to complexity theory   The existence of a simulation $d \to e$ seems to be conceptually analogous to the existence of a reduction from one problem another. How far can one push this analogy to complexity theory into "contextuality theory"? For example, are there complete problems for a "contextuality-class", at least if one fixes the measurement scenario?

Stronger notions of morphisms   One might also wonder if stronger notions of a morphism are warranted. For example, if the students are allowed to share quantum resources, one should get a notion of a "quantum simulation". The R-stochastic simulations use the Kleisli category of the distribution monad. Maybe the notion of a quantum reduction uses something like the quantum monad from [4].

Perhaps the most pressing issue is preprocessing: intuitively, when using $d\colon \langle Y, \mathcal{N}, P \rangle$ to simulate $e\colon \langle X, \mathcal{M}, O \rangle$, one should, for a given $x \in X$, be able to use convex mixtures of measurements in Y. This might be needed to fully capture the meaning of "d can be used to simulate e given classical shared correlations". Indeed, preprocessing is a fundamental part of the maps in [8], and similarly they are routinely used in [9] and [10]. In [10] one considers several copies of PR-boxes and the correlations one can build from them, allowing for simulations where the output of one PR box is fed into another. These two features - measurement choices depending on classical randomness or on results of other measurements - result in wider notion of simulation, and we leave formalizing it in the sheaf-theoretic framework as an open question.

One possibility might be to think of a "simulation $d \to e$ with preprocessing" as a ordinary simulation $d' \to e$, where d' is built from d and the randomization needed in the preprocessing - in a sense, d' is the sequential composition of d with a non-contextual model. For this approach to work, one probably needs a good theory of sequential composition of empirical models. The alternative approach is to bite the bullet and let $\pi\colon X \to Y$ be stochastic instead of deterministic. The worry of course being that either the theory does not work or becomes too unwieldy to be useful. In particular, combining both a relation $\pi X \to Y$ and preprocessing might lead to trouble: for instance, what does it mean to mix a measurement x_1 with a pair of measurements x_2 and x_3? Incorporating both of these aspects might be difficult, perhaps for the same reasons combining nondeterministic and probabilistic computing is difficult [21].

Acknowledgements

References

[1] Samson Abramsky (2013): *Relational hidden variables and non-locality*. *Studia Logica* 101(2), pp. 411–452.

[2] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal & Shane Mansfield (2015): *Contextuality, Cohomology and Paradox*. In: *Proceedings of CSL 2015, Leibniz International Proceedings in Informatics (LIPIcs)* 41, pp. 211–228.

[3] Samson Abramsky, Rui Soares Barbosa & Shane Mansfield (2017): *Contextual fraction as a measure of contextuality*. *Physical review letters* 119(5), p. 050504.

[4] Samson Abramsky, Rui Soares Barbosa, Nadish de Silva & Octavio Zapata (2017): *The quantum monad on relational structures*. *arXiv preprint arXiv:1705.07310*.

[5] Samson Abramsky & Adam Brandenburger (2011): *The sheaf-theoretic structure of non-locality and contextuality*. *New Journal of Physics* 13(11), p. 113036.

[6] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2011): *The Cohomology of Non-Locality and Contextuality*. In: *Proceedings of QPL 2011*, Electronic Proceedings in Theoretical Computer Science, Citeseer, pp. 1–14.

[7] Antonio Acín, Tobias Fritz, Anthony Leverrier & Ana Belén Sainz (2015): *A combinatorial approach to nonlocality and contextuality*. *Communications in Mathematical Physics* 334(2), pp. 533–628.

[8] Barbara Amaral, Adán Cabello, Marcelo Terra Cunha & Leandro Aolita (2017): *Noncontextual wirings*. *arXiv preprint arXiv:1705.07911*.

[9] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu & David Roberts (2005): *Nonlocal correlations as an information-theoretic resource*. *Physical Review A* 71(2), p. 022101.

[10] Jonathan Barrett & Stefano Pironio (2005): *Popescu-Rohrlich correlations as a unit of nonlocality*. *Physical review letters* 95(14), p. 140401.

[11] Francis Borceux (1994): *Handbook of Categorical Algebra 2: Categories and Structures (Encyclopedia of Mathematics and Its Applications)*. Cambridge University Press.

[12] Bob Coecke, Tobias Fritz & Robert W Spekkens (2016): *A mathematical theory of resources*. *Information and Computation* 250, pp. 59–86.

[13] Oscar Cunningham & Chris Heunen (2017): *Purity through factorisation*. In: *Proceedings of QPL 2017, Electronic Proceedings in Theoretical Computer Science* 266.

[14] Cristhiano Duarte & Barbara Amaral (2017): *Resource theory of contextuality for arbitrary prepare-and-measure experiments*. *arXiv preprint arXiv:1711.10465*.

[15] Bart Jacobs (2010): *Convexity, duality and effects*. In: *IFIP International Conference on Theoretical Computer Science*, Springer, pp. 1–19.

[16] Saunders MacLane & Ieke Moerdijk (1992): *Sheaves in geometry and logic: A first introduction to topos theory*. Springer-Verlag.

[17] Shane Mansfield (2013): *The Mathematical Structure of Non- locality and Contextuality*. D. Phil. thesis Oxford University.

[18] Shane Mansfield & Tobias Fritz (2012): *Hardys non-locality paradox and possibilistic conditions for non-locality*. *Foundations of Physics* 42(5), pp. 709–719.

[19] Robert W Spekkens (2005): *Contextuality for preparations, transformations, and unsharp measurements*. *Physical Review A* 71(5), p. 052108.

[20] Sam Staton & Sander Uijlen (2015): *Effect algebras, presheaves, non-locality and contextuality*. In: *International Colloquium on Automata, Languages, and Programming*, Springer, pp. 401–413.

[21] Daniele Varacca & Glynn Winskel (2006): *Distributing probability over non-determinism*. *Mathematical Structures in Computer Science* 16(1), p. 87113, doi:10.1017/S0960129505005074.

[22] Nikolai Nikolaevich Vorobev (1962): *Consistent families of measures and their extensions*. *Theory of Probability & Its Applications* 7(2), pp. 147–163.

[23] Linde Wester (2017): *Almost Equivalent Paradigms of Contextuality*. In: *Proceedings of QPL 2017, Electronic Proceedings in Theoretical Computer Science* 266.

Appendix

Proof of Theorem 4.7. We define a simulation $e|_{X\setminus\{x\}} \to e$. First of all, let $S \in \mathscr{M}$ be the only maximal face containing x. We define a simplicial relation $\pi\colon \langle X, \mathscr{M}\rangle \to \langle X\setminus\{x\}, \{C\setminus\{x\}|C \in \mathscr{M}\}\rangle$ by

$$\pi(y) = \begin{cases} C\setminus\{x\} & \text{if } y = x \\ \{y\} & \text{otherwise} \end{cases}$$

Next we define a natural transformation in parts using Lemma 3.12. For that purpose, we partition X into $U_1 := \{x\}$ and $U_2 := X\setminus\{x\}$. On U_2 the relation π restricts to the identity, so we can define $\sigma^2 = \eta$, i.e. the identity for the Kleisli composition. For the other half, we just need to define a map $\sigma_x\colon O^{\pi(x)} = O^{C\setminus\{x\}} \to D_R(O^x)$. Fix $o \in O$. Given $s\colon C\setminus X \to O$, we set

$$\sigma_x(s) = \begin{cases} e_C(-|s) & \text{if } s \in \text{supp}_{C\setminus\{x\}}(e_C) \\ 1\cdot o & \text{otherwise.} \end{cases}$$

To check that the resulting morphism $(\pi, \sigma \otimes \eta)$ of measurement scenarios defines a simulation, $e|_{X\setminus\{x\}} \to e$, consider first a context $D \in \mathscr{M}$ that does not contain x. Then $D \subset U_2$ and $\sigma \otimes \eta$ restricts to η, which is the identity for the Kleisli composition.

Consider now the context C containing the deleted point x. Now $\pi(C) = C\setminus\{x\}$, and write $e|_{C\setminus\{x\}} =: d$ in the form $\sum_{s\in\text{supp}\,d} d(s)s$. Then

$$\mu D_R(\sigma\otimes\eta)_C d = \mu D_R(\sigma\otimes\eta)_C \sum_{s\in\text{supp}\,d} d(s)s$$

$$= \mu \sum_{s\in\text{supp}\,d} d(s)(\sigma\otimes\eta)_C(s)$$

$$= \mu \sum_{s\in\text{supp}\,d} d(s)(\sigma_x(s)\otimes\eta_{C\setminus\{x\}}(s))$$

$$= \mu \sum_{s\in\text{supp}\,d} d(s)(e|_C(-|s)\otimes 1s)$$

This being equal to e_C boils down to the fact that $p(x,y) = p(x|y)p(y)$ for conditional R-distributions.

$\square$