# Functorial Aggregation

David I. Spivak

TOPOS INSTITUTE

Workshop on Polynomial Functors
2022 March 18

# Outline

# Why think about databases?

I'm interested in sense-making. How do we make sense of the world?

- We're here together, each with our own purpose and abilities.
- We're engaged in the activity of collective sense-making.
- I'm want to spread a *sense* of how **Poly** relates to information.

Imagine that sense is "contained" somewhere and that it can be transferred.

- If our ability to deal effectively with the world were contained...
- ...in our brain, then we could ask "what's the brain's data structure?"
- And if our data structures are different, then how is info transferred?
- Are you getting this? If so, what's the story of *how that works*?

I think of mathematical fields as accounting systems.

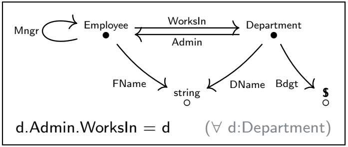
- Arithmetic accounts for the flow of quantities, as in finance.
- Hilbert spaces account for the states of elementary particles, as in QM.
- Probability distributions account for likelihoods, as in game theory.
- What's a good accounting system for how we collectively make sense?

# Categorical databases

When I first started out on this question, I began with databases.

- Their mundane and humble but widely-used and easily conceptualized.
- The way I conceptualized them was as copresheaves $F\colon \mathcal{C} \to \mathbf{Set}$.
- The site $\mathcal{C}$ is called the *schema* and the copresheaf is the *instance*.

mySchema :=




| Employee | FName | WorksIn | Mngr |
|---|---|---|---|
| 1 | Alan | 101 | 2 |
| 2 | Ruth | 101 | 2 |
| 3 | Kris | 102 | 3 |

| String |
|---|
| Alan |
| IT |
| ⋮ |

| Department | DName | Admin | Bdgt |
|---|---|---|---|
| 101 | Sales | 1 | $10 |
| 102 | IT | 3 | $5 |

| x : $ |
|---|
| $5 |
| $6 |
| ⋮ |

- The schema holds the *structure* of your knowledge...
- ...and the instance holds all your *examples* within that structure.

For those who don't care about databases, this talk is about copresheaves.

# Querying and aggregating

The two most common thing to do with databases is *query and aggregate*.

- Querying a database means performing a limit operation.
  - Example: find all pairs of people with the same favorite book.

$$\begin{tikzcd} Q \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & P \arrow[d, "f"] \\ P \arrow[r, "f"'] & B \end{tikzcd}$$

  - This is called a *conjunctive query*, "an $X$ and a $Y$ where..."
  - More generally a *disjoint union of conj'ive queries (duc-query)*...
  - ... is a coproduct of these, e.g. same favorite book or movie.
- Aggregating is "integrating along compact fibers".
  - Assign to each $b \in B$ the number of people whose fave book is $b$.

$$\begin{tikzcd} P \arrow[r, "s"] \arrow[d, "f"'] & \mathbb{R} \\ B \arrow[ur, dashed, "(\text{sum } s)_f"'] & \end{tikzcd}$$

  - Or assign to $b$ the total salary of everyone whose fave is $b$.
  - Rather than searching, this is summarizing, reporting.

# Beauty and the beast

Both querying and aggregating are crucial, but one is cat'ly better behaved.

- Querying is part of a larger story called *data migration*.
- Given two categories (DB schemas) $\mathcal{C}, \mathcal{D}$, a data migration functor...
- ... $\mathcal{C} \mathrel{\triangleleft\!\!-\!\!\!-\!\!\!-\!\!\triangleleft} \mathcal{D}$ is a parametric right adjoint $\mathcal{D}$-**Set** $\to \mathcal{C}$-**Set**.
- These have nice characterizations, some of which we'll discuss, and...
- ... have been implemented in open-source categorically-minded code.

In contrast, aggregation has seemingly not received a categ'al formulation.

- It's not even clear what sort of properties are desirable.
- For example, aggregation is not natural wrt. copresheaf morphisms.
- Consider: what is preserved by a commutative square $f' \to f$?

$$\begin{tikzcd} P' \arrow[r] \arrow[d, "f'"'] & P \arrow[r, "s"] \arrow[d, "f"'] & \mathbb{R} \\ B' \arrow[r] & B \arrow[ur, dashed, "(\mathtt{sum}\ s)_f"'] & \end{tikzcd}$$

Can you make a match between beauty and the beast? "Be my guest!"

# Poly-amory: could one category be enough?

I've lately been totally enamored with **Poly** and its comonoids $\mathbb{C}\mathbf{at}^\sharp$.

- First, it is an excellent setting for thinking about things I care about.
  - It is a natural setting for interacting dynamical systems.
  - And thanks to the results of Ahman-Uustalu and Garner...
  - ...it is a natural setting for databases and data migration.
  - Both of these seem relevant when accounting for sense-making.
- Second, **Poly** has loads and loads of structure.
  - Coproducts and products that agree with usual polynomial arithmetic;
  - All limits and colimits;
  - At least three orthogonal factorization systems;
  - A symmetric monoidal structure $\otimes$ distributing over $+$;
  - A cartesian closure $q^p$ and monoidal closure $[p, q]$ for $\otimes$;
  - Another nonsymmetric monoidal structure $\triangleleft$ that's duoidal with $\otimes$;
  - A left $\triangleleft$-coclosure $\begin{bmatrix} - \\ - \end{bmatrix}$, meaning $\mathbf{Poly}(p, q \triangleleft r) \cong \mathbf{Poly}(\begin{bmatrix} r \\ p \end{bmatrix}, q)$;
  - An indexed right $\triangleleft$-coclosure (Myers?), i.e. $\mathbf{Poly}(p, q \triangleleft r) \cong \sum_{f\colon\ p(1) \to q(1)} \mathbf{Poly}(p \overset{f}{\frown} q, r)$;
  - An indexed right $\otimes$-coclosure (Niu?), i.e. $\mathbf{Poly}(p, q \otimes r) \cong \sum_{f\colon\ p(1) \to q(1)} \mathbf{Poly}(p \overset{f}{\nearrow} q, r)$;
  - At least eight monoidal structures in total;
  - $\triangleleft$-monoids generalize plain operads;
  - $\triangleleft$-comonoids are exactly categories; bicomodules are data migrations. This is $\mathbb{C}\mathbf{at}^\sharp$.
  - For the above and more, see "A reference for categorical structures on **Poly**", arXiv: 2202.00534

It was love at first sight. I'm committed to solving problems as a team.

# Aggregation poses a "purity of methods" problem

"Solving aggregation" is not well-defined.

- Given a map $E \to B$ and a map $E \to \mathbb{R}$, you "just integrate".
- It's hard to know what problem needs solving.

According to Detlefsen & Arana, "purity of methods" has a long tradition.

- Aristotle, Newton, Lagrange, Gauss, Bolzano, Frege,... all sought it.
- Erdös wanted a non-$\mathbb{C}$ proof of Hadamard's prime number thm ($\frac{n}{\log n}$).
- Bolzano phrased it as searching for "a thorough way of thinking."

So this suggests a ways forward.

- Since **Poly** is great for thinking about data migration (as I'll discuss)...
- ...it is a "purity of methods" issue to get aggregation into **Poly** as well.
- So the goal is to give an account of aggregation using...
- ...only monoidal/universal structures available in the **Poly** ecosystem.
- Pursuing it led to several new structures, which I'll tell you about.

# Plan of the talk

Now that I've introduced the topic, here's the plan for my remaining time.

- Give background on **Poly**, its monoidal closure and mon'l coclosure.
- Discuss $\mathbb{C}\mathbf{at}^\sharp = \mathbb{C}\mathbf{omon}(\mathbf{Poly})$, natural home of data migration.
- Show how aggregation-useful structures on **Poly** generalize to $\mathbb{C}\mathbf{at}^\sharp$.
- Explain aggregation with the structures we've explored.
- Conclude with a summary.

# Outline

# Poly: coproducts of representables $\mathbf{Set} \to \mathbf{Set}$

A polynomial functor is a coproduct of representables $\mathbf{Set} \to \mathbf{Set}$:

- For any set $E$, denote the functor it represents by $y^E := \mathbf{Set}(E, -)$.
- E.g. $y = y^1$ is identity, $y^0 = 1$ is constant, and $y^E(1) \cong 1$ for any $E$.
- A *polynomial* is a disjoint union of representables $p \cong \sum_{b \in B} y^{E_b}$.
- Note that $p(1) \cong B$ so, we can denote polynomials as follows:

$$p := \sum_{I \in p(1)} y^{p[I]}$$

Morphisms $p \xrightarrow{\varphi} q$ are just natural transformations $\begin{tikzcd} \mathbf{Set} \arrow[r, bend left, "p"] \arrow[r, bend right, "q"'] \arrow[r, phantom, "\Downarrow\varphi"] & \mathbf{Set} \end{tikzcd}$

- Combinatorially, a map $\varphi \colon p \to q$ can be given in two parts:
- A function $\varphi_1 \colon p(1) \to q(1)$ "forward on positions" and...
- ...for each $I \in p(1)$, a function $q[\varphi_1 I] \to p[I]$ "backward on directions"

A polynomial can be viewed as a functor or just as a combinatorial object.

- Polynomials can be viewed as functors; this is like "querying".
- The functor $p$ "migrates data", sending $X \in \mathbf{Set}$ to $p(X) \in \mathbf{Set}$.
- But it's often helpful just to think of $p$ as a data structure.

# Dirichlet product $\otimes$ and its closure

**Poly** admits many monoidal structures, e.g. coproduct and product $(+, \times)$.

- Among the most useful is Dirichlet product $\otimes$; its unit is $y$.
- If you think of a polynomial $p$ as a bundle, $\left(\sum_{I \in p(1)} p[I]\right) \to p(1)$...
- ...then $p \otimes q$ is just product of base and total spaces for $p, q \in \mathbf{Poly}$.

$$p \otimes q \cong \sum_{(I,J) \in p(1) \times q(1)} y^{p[I] \times q[J]}$$

The $\otimes$-structure has a closure: $\mathbf{Poly}(p \otimes q, r) \cong \mathbf{Poly}(p, [q, r])$:

$$[q, r] \cong \sum_{\varphi \in \mathbf{Poly}(q,r)} y^{\sum_{J \in q(1)} r[\varphi_1 J]}$$

Like so much in **Poly**, both $\otimes$ and $[-, -]$ generalize to $\mathbb{C}\mathbf{at}^\sharp$.

- The polynomial $y$ is a *dualizing object*: for any $A \in \mathbf{Set}$...
- ...we have isomorphisms $[Ay, y] \cong y^A$ and $[y^A, y] \cong Ay$.
- We write $\overline{Ay} \cong y^A$. This generalizes to an important part of our story.

# Substitution product $\triangleleft$ and its coclosure

The other important monoidal product is called *substitution* or *composition*.

- The composite $p \triangleleft q := p \circ q$ of polynomial functors is a polynomial.
- E.g. if $p = y^2$ and $q = y + 1$, then $p \triangleleft q \cong y^2 + 2y + 1$.
- (There are many reasons for $\triangleleft$ instead of $\circ$. One is that we want to reserve $\circ$ for morphisms; $\triangleleft$ is "composing" objects!)
- This monoidal product $\triangleleft$ preserves equalizers in both variables.
- It and $\otimes$ are duoidal: $(p_1 \triangleleft p_2) \otimes (q_1 \triangleleft q_2) \to (p_1 \otimes q_1) \triangleleft (p_2 \otimes q_2)$.

In fact, $\triangleleft$ has a right coclosure (Myers?) and an indexed left coclosure.

- We'll only need the former, $\mathbf{Poly}(p, q \triangleleft p') \cong \mathbf{Poly}\left(\begin{bmatrix} p' \\ p \end{bmatrix}, q\right)$.
- For any $p \in \mathbf{Poly}$ the polynomial $\begin{bmatrix} p \\ p \end{bmatrix}$ is a $\triangleleft$-comonoid (Meyers).
- We'll see that $\triangleleft$-comonoids are categories; which one is this?
- It's the *full internal subcat'y of* $\mathbf{Set}^{\mathrm{op}}$ (Jacobs) spanned by $p$-fibers.

We'll rely heavily on this in a special case: $u = \mathsf{List} = \sum_{N \in \mathbb{N}} y^N$.

- Then $\begin{bmatrix} u \\ u \end{bmatrix}$ is a skeleton of $\mathbf{Fin}^{\mathrm{op}}$.
- Later we'll get its opposite as the dual, $\overline{\begin{bmatrix} u \\ u \end{bmatrix}} \simeq \mathbf{Fin}$.

# Outline

1 Introduction

2 Background on Poly

3 $\mathbb{C}\mathrm{at}^\sharp$, home of data migration
- Shulman, Ahman-Uustalu, Garner
- Databases for intuition
- Categorical structures

4 Aggregation

5 Conclusion

# $\mathbb{C}\mathbf{at}^\sharp := \mathbb{C}\mathbf{omon}(\mathbf{Poly})$

Shulman: if equipment $\mathbb{P}$ has good equalizers, $\mathbb{C}\mathbf{omon}(\mathbb{P})$ is an equipment.

- An *equipment* is a kind of double cat'y, where 2-cells can be "cartesian."
- Any bicat'y—and hence any monoidal category—is one, called *globular*.
- So $(\mathbf{Poly}, y, \triangleleft)$ is a equipment! It happens to be vertically-trivial.
- We said **Poly** has "good equalizers": i.e. they are preserved by $\triangleleft$.
- Shulman tells us that $\mathbb{C}\mathbf{at}^\sharp := \mathbb{C}\mathbf{omon}(\mathbf{Poly})$ is also an equipment.

Ahman-Uustalu: the objects of $\mathbb{C}\mathbf{omon}(\mathbf{Poly})$ are categories.

- A comonoid in $(\mathbf{Poly}, y, \triangleleft)$ consists of $(c, \epsilon, \delta)$ where $c \in \mathbf{Poly}$ and...
- ...$\epsilon\colon c \to y$ and $\delta\colon c \to c \triangleleft c$ are (counital & coassoc'tive) maps.
- This turns out to force $c \cong \sum_{a \in \mathrm{Ob}(\mathcal{C})} y^{\mathcal{C}[a]}$ for some category $\mathcal{C}$,...
- ...where $\mathcal{C}[a] := \sum_{a' \in \mathrm{Ob}(\mathcal{C})} \mathrm{Hom}_{\mathcal{C}}(a, a')$. Say $c$ is $\mathcal{C}$'s *outfacing poly'l*.
- Then $\epsilon$ gives identities and $\delta$ gives codomains and composition.

Garner: horizontal morphisms of $\mathbb{C}\mathbf{omon}(\mathbf{Poly})$ are data migrations.

- He didn't say it that way. He said that a bicomodule $\mathcal{C} \overset{p}{\triangleleft\!\!\longrightarrow\!\!\triangleleft} \mathcal{D}$...
- ...is a parametric right adjoint $\mathbf{Set}^{\mathcal{D}} \to \mathbf{Set}^{\mathcal{C}}$. I'll explain soon.
- I denote the cat'y of $(c, d)$-bicomodules by $c$-$\mathbf{Set}[d]$.

# $\mathbb{P}$olyFun and $\mathbb{S}$pan as subequipments of $\mathbb{C}\mathbf{at}^\sharp$

To understand $\mathbb{C}\mathbf{at}^\sharp := \mathbb{C}\mathbf{omon}(\mathbf{Poly})$, let's start with something familiar.

- Gambino-Kock showed that sets, functions, and multi-variate poly's...
- ...form an equipment, called $\mathbb{P}\mathbf{olyFun}$. (And similarly for arbitrary LCC cat'y in place of **Set**).
- It sits inside $\mathbb{C}\mathbf{at}^\sharp$ as the full subequipment spanned by discrete caty's.
- Discrete caty's are those whose outfacing poly'l is linear, $I\mathcal{y}$ for $I \in \mathbf{Set}$.

$$\begin{tikzcd}
I \arrow[dd] & E \arrow[l] \arrow[r] & B \arrow[r] \arrow[d, equal] & J \arrow[dd] \\
 & E' \times_{B'} B \arrow[u, "\varphi^\sharp"] \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner" very near start] & B \arrow[d, "\varphi_1"] & \\
I' & E' \arrow[l] \arrow[r] & B' \arrow[r] & J'
\end{tikzcd}
\quad \leftrightsquigarrow \quad
\begin{tikzcd}
I\mathcal{y} \arrow[d] & J\mathcal{y} \arrow[l, "p"'] \arrow[d] \arrow[l, phantom, shift left=6, ""] \\
I'\mathcal{y} & J'\mathcal{y} \arrow[l, "p'"] \arrow[ul, phantom, "\Downarrow \varphi"]
\end{tikzcd}$$

All the "activity" is subsumed under the def'n of comonoid, comodule.

The usual double cat'y of spans sits inside: $\mathbb{S}\mathbf{pan} \subseteq \mathbb{P}\mathbf{olyFun} \subseteq \mathbb{C}\mathbf{at}^\sharp$.

- $\mathbb{S}\mathbf{pan} \subseteq \mathbb{C}\mathbf{at}^\sharp$ is the subequipment where every poly is linear!
- Discrete carrier makes: a cat'y be discrete, a bicomodule be a span.

# Database schemas and Duc-queries

We can get more intuition for $\mathbb{C}\mathbf{at}^\sharp$ by thinking about databases.

- The indexing cat'y for graphs is $\mathcal{G} := \boxed{\bullet \rightrightarrows \bullet}$, carried by $g := y^3 + y$.
- Bicomodules $c \triangleleft\!\xrightarrow{X}\!\triangleleft\ 0$ can be ident'd with copresheaves $c \xrightarrow{X} \mathbf{Set}$.
- So a graph is just a bicomodule $g \triangleleft\!\xrightarrow{X}\!\triangleleft\ 0$. Call $X$ a $g$-set or $\mathcal{G}$-set.
- Think of $\mathcal{G}$ as a database *schema*, an arrangement for sets, and...
- ...think of $X\colon \mathcal{G} \to \mathbf{Set}$ as a $\mathcal{G}$-*instance*, some sets so-arranged.

We can move data between database schemas using duc-queries.

- The cat'y of *conjunctive queries on* $\mathcal{C}$ is $(\mathcal{C}\text{-}\mathbf{Set})^{\mathrm{op}}$. Idea:
- For $Q \in \mathcal{C}\text{-}\mathbf{Set}$ we have a functor $\mathcal{C}\text{-}\mathbf{Set}(Q, -)\colon \mathcal{C}\text{-}\mathbf{Set} \to \mathbf{Set}$.
- We think of $Q$ as a query: find me all $Q$-shapes in $-$. Contravariant in $Q$.
- E.g. there's a graph $Q_n$ for which $\mathcal{G}\text{-}\mathbf{Set}(Q_n, -)$ returns length-$n$ paths.
- If we want all paths, we need a disjoint union of conjunctive query's.

A data migration (bicomod) $\mathcal{C} \triangleleft\!\xrightarrow{p}\!\triangleleft\ \mathcal{D}$ is a $\mathcal{C}$-indexed duc-query on $\mathcal{D}$.

- It functorially assigns a duc-query on $\mathcal{D}$ to each $c \in \mathcal{C}$.
- Given a $\mathcal{D}$-instance $\mathcal{D} \triangleleft\!\longrightarrow\!\triangleleft\ 0$, composition returns a $\mathcal{C}$-instance.

# Adjoint prafunctors

Data migrations = Bicomodules = *parametric right adjoint* functors.

- A bicomodule $c \xleftarrow{p} d$ is a prafunctor $d$-**Set** $\to$ $c$-**Set**.
- It is also a $c$-indexed duc-query on $d$. This can be helpful.
- E.g., a profunctor $c^{\mathsf{op}} \times d \to \mathbf{Set}$ is a special case of prafunctor.
- It can be identified with a $c$-indexed *conjunctive* query on $d$, no sums.

When is a prafunctor $c \xleftarrow{p} d$ a right adjoint in $\mathbb{C}\mathbf{at}^\sharp$?

- Two conditions: it is a right adjoint functor $d$-**Set** $\to$ $c$-**Set** and...
- ...the associated left adjoint $c$-**Set** $\to$ $d$-**Set** is also a prafunctor.
- For example for any functor $F\colon c \to d$, we have $\Delta_F \dashv \Pi_F$.
- For $c \xleftarrow{p} d$, I denote its left adjoint by $d \xleftarrow{p^\dagger} c$.

In the subcategory $\mathbb{P}\mathbf{olyFun} = \mathbb{C}\mathbf{at}^\sharp_{\text{disc}}$, adjoints are easy to characterize:

- Left adjoints are those $p$ with linear carrier (spans), and...
- ...right adjoints are the profunctors, i.e. $c$-indexed conjunctive queries.

# External and internal $\otimes$

The equipment $\mathbb{C}\mathbf{at}^\sharp$ has lots of structure, e.g. it is monoidal.

- There is a double functor $\otimes\colon \mathbb{C}\mathbf{at}^\sharp \times \mathbb{C}\mathbf{at}^\sharp \to \mathbb{C}\mathbf{at}^\sharp$.
- It is an (external) symmetric monoidal structure on $\mathbb{C}\mathbf{at}^\sharp$.
- On objects, $c \otimes d$ is the usual product of categories.

It also has local $\otimes$-monoidal structures.

- For any $c, d \in \mathbb{C}\mathbf{at}^\sharp$ the cat'y $c$-$\mathbf{Set}[d]$ has an induced $\otimes$-structure.
- That is, for any two bicomodules $c \xleftarrow{p,q} d$, there is...
- ...a bicomodule $c \xleftarrow{p\ {}_c\otimes_d\ q} d$. The local unit is $c(1)y^{d(1)}$.
- These fit together "duoidally":

$$
\begin{tikzcd}
c_0 & c_1 \arrow[l, shift right, "p_1"'] \arrow[l, shift left, "q_1"] & c_2 \arrow[l, shift right, "p_2"'] \arrow[l, shift left, "q_2"]
\end{tikzcd}
\quad\rightsquigarrow\quad
\begin{tikzcd}[column sep=huge]
c_0 & c_2 \arrow[l, bend right, "(p_1 \lhd_{c_1} p_2)\ {}_{c_0}\otimes_{c_2}\ (q_1 \lhd_{c_1} q_2)"', ""{name=U, below}] \arrow[l, bend left, "(p_1\ {}_{c_0}\otimes_{c_1}\ q_1) \lhd_{c_1} (p_2\ {}_{c_1}\otimes_{c_2}\ q_2)", ""{name=D, above}] \arrow[Rightarrow, from=U, to=D, "\text{duoid}"]
\end{tikzcd}
$$

# Local closures and dualizing object

The local $\otimes$-structures have closures.

- That is, for $p, q, r \in c\text{-}\mathbf{Set}[d]$, there is a natural isomorphism

$$c\text{-}\mathbf{Set}[d](p\ {}_c\!\otimes_d\ q\ ,\ r) \cong c\text{-}\mathbf{Set}[d](p\ ,\ {}_c[q, r]_d)$$

- Wanted: other equip's with local (duoidal) monoidal-closed structure.

For any sets $C, D$, there's a dualizing object in $C\text{-}\mathbf{Set}[D]{=}\mathbf{Poly}_{GK}(C, D)$.

- It's the terminal span, $C \leftarrow (C \times D) \rightarrow D$, i.e. $\mathcal{C}y \xleftarrow{\mathcal{CD}y} \mathcal{D}y$.
- Calling it $\perp := \mathcal{CD}y$, the functor $\overline{\cdot} := {}_C[-, \perp]_D$ provides a duality...
- If $p \in C\text{-}\mathbf{Set}[D]$ is linear then $[p, \perp]$ is conjunctive, and vice versa.
- In particular $\overline{\overline{p}} \cong p$ for any linear or conjunctive $p$.
- It generalizes $[\mathcal{A}y, y] \cong y^A$ from earlier.

# Transposing a span, "oppositing" a category

The idea of $\overline{p}$ is that it transforms $\Sigma_F \circ \Delta_G$ into $\Pi_F \circ \Delta_G$.

- That's not its adjoint!
- The adjoint of $\Sigma_F \circ \Delta_G$ is $\Pi_G \circ \Delta_F$.
- So what is the adjoint of the dual or the dual of the adjoint?
- Answer: $\Sigma_F \circ \Delta_G \;\mapsto\; \Pi_F \circ \Delta_G \;\mapsto\; \Sigma_G \circ \Delta_F$. (The other works too.)

On the level of spans, this is the transpose!

- The transpose operation is a composite of two more primitive ones.
- This doesn't happen within $\mathbb{S}$**pan**; kind of like contour integrals.

Similarly, the opposite of a category is a composite of two operations.

- A category $\mathcal{C}$ can be viewed as a monad in $\mathbb{S}$**pan**, and its adjoint is...
- ...a comonoid $c(1)y \xleftarrow{c} c(1)y$, which is a cat'y in a different way!
- And the dual of that comonoid is again a monad in $\mathbb{S}$**pan**, namely $\mathcal{C}^{\mathsf{op}}$.
- In particular, with $u = \sum_{N \in \mathbb{N}} y^N$, we will use (twice) that $\overline{\begin{bmatrix} u \\ u \end{bmatrix}} \simeq$ **Fin**.

# Outline

# Finitary instances

For $X\colon c \to \mathbf{Set}$, i.e. $c \triangleleft\!\xleftarrow{X}\!\triangleleft\ 0$, the following are equivalent

- the copresheaf $X$ is *finitary*, i.e. it factors through $\mathbf{Fin} \subseteq \mathbf{Set}$.
- there exists a function $\ulcorner X \urcorner\colon c(1) \to u(1)$ with

$$\begin{tikzcd} c(1)y \arrow[d, "\ulcorner X \urcorner"'] \arrow[dr, phantom, "\text{cart}"] & 0 \arrow[l, "X"'] \arrow[d, equal] \\ u(1)y & 0 \arrow[l, "\overline{u}"] \end{tikzcd}$$

- There exists a monad map $\ulcorner X \urcorner_1$ as shown here:

$$\begin{tikzcd} c(1)y \arrow[d, "\ulcorner X \urcorner"'] \arrow[dr, phantom, "\Downarrow \ulcorner X \urcorner_1"] & c(1)y \arrow[l, "c^{\downarrow}"'] \arrow[d, "\ulcorner X \urcorner"] \\ u(1)y & u(1)y \arrow[l, "\overline{\begin{bmatrix} u \\ u \end{bmatrix}}"] \end{tikzcd} \qquad c^{\downarrow} = \sum_{a \in c(1)} c^{\text{op}}[a]y.$$

  for which the $c^{\downarrow}$-algebra induced by $\overline{u}$ is $X$.

I know that's impossible to follow; sorry! The gist: "everything works!"

# Commutative monoids as Fin-algebras

A database schema assigns a comm've monoid $(M_a, \circledast_a)$ to each $a \in c(1)$.

- Assigning the set $M_a$ to each $a \in c(1)$ is a bicomodule $y \xleftarrow{My} Ay$.
- Consider the following diagram that coerces $\circledast$ into the picture:

$$\begin{tikzcd}
 & u(1)y \arrow[dl, "\overline{\begin{bmatrix} u \\ u \end{bmatrix}}"'] & & y \arrow[ll, "u"'] \arrow[d, phantom, "\circledast\Downarrow"] & \\
u(1)y & & & y \arrow[lll, "u"] & c(1)y \arrow[l, "My"] \arrow[ul, "My"']
\end{tikzcd}$$

- The composite $u(1) \xleftarrow{u} y \xleftarrow{My} c(1)y$ assigns...
- ... to each $N \in \mathbb{N}$ and $a \in c(1)$ the set $(M_a)^N$.

So what does the 2-cell say, and what does being a $\overline{\begin{bmatrix} u \\ u \end{bmatrix}}$-module mean?

- Given a function $f \colon N \to N'$, an object $a \in c(1)$, and $m \in (M_a)^N$...
- ...there is an induced $(\circledast m)_f \in (M_a)^{N'}$. Integration along fibers.
- $u \triangleleft My$ being a $\overline{\begin{bmatrix} u \\ u \end{bmatrix}}$-algebra means it works with ids and composites.

# Aggregation

The thing we've worked so hard for is as follows.

- Suppose we have a category $c$ and a copresheaf $X\colon c \to \mathbf{Set}$ and...
- ...a commutative monoid $M_a$ and a map $s_a\colon X_a \to M_a$ for each $a \in c(1)$.
- Then given $f\colon a \to b$ in $c$, and given $y \in X(b)$, we want:
- ... to take the fiber $\{x \in X(a) \mid x.f = y\}$ and "add 'em up".
- That is, take $\circledast_{\{x \mid x.f=y\}} s_a(x)$.

$$
\begin{tikzcd}
X_a \arrow[r, "s_a"] \arrow[d, "X_f"'] & M_a \\
X_b \arrow[ur, dashed, "(\circledast\, s_a)_f"']
\end{tikzcd}
$$

We have accomplished this now, using pieces we've collected.

$$
\begin{tikzcd}
c(1)y \arrow[d, "\ulcorner X \urcorner"'] \arrow[r, phantom, "\Downarrow \ulcorner X \urcorner_1"] & c(1)y \arrow[l, "c\dagger"'] \arrow[d, "\ulcorner X \urcorner"] \arrow[rr, phantom, "\Downarrow s"] & & c(1)y \arrow[ll, "c(1)y"'] \arrow[d, equal] \arrow[lll, bend right, "c\dagger"', ""{name=T}] \\
u(1)y & u(1)y \arrow[l, "\overline{\begin{bmatrix} u \\ u \end{bmatrix}}"] & y \arrow[l, "u"] \arrow[d, phantom, "\Downarrow\circledast"] & c(1)y \arrow[l, "My"] \arrow[dl, bend left, "My"] \\
& & y \arrow[ull, bend left, "u"] &
\end{tikzcd}
$$

# Outline

# Summary

Aggregation is of central importance in database practice.

- Add up salaries, count things, collect each fiber into a set, etc.
- If we also have "calculated fields" (not too hard), you can...
- ... take averages, plot graphs, etc. Aggregation is very powerful.

There's a really nice categorical story for data migration.

- It is that $\mathbb{C}\mathbf{at}^\sharp = \mathbb{C}\mathbf{omon}(\mathbf{Poly})$ is categories and prafunctors.
- And prafunctors are data migrations (e.g. find all paths in a graph).
- But a categorical formulation of aggregation has been missing.

But **Poly** is so highly-structured, we asked if it might include aggregation.

- Using adjoint prafunctors, local monoidal closures, and coclosures...
- ...we found a way to say what we needed to say.
- It's not as plain and simple as I'd like, but there's likely a better way.
- We tested the mettle of **Poly** and it was indeed up to the task!

*Thank you for your time; questions and comments welcome!*