# Protocol Labs Research ∩ ACT

David A. Dalrymple
@davidad

Applied Category Theory 2022 (ACT2022), Glasgow

2022-07-21

# Who are Protocol Labs?

- We're best known for:
  - IPFS (InterPlanetary File System): millions of unique weekly active users
  - Filecoin: 17 exabytes of active storage, market cap >£1 billion (even now)
  - `libp2p`: core abstractions of IPFS & Filecoin, also used by Ethereum 2.0
- Mission: **breakthroughs in computing to drive humanity forward**
- Unusual conjunction of techno-optimism, tech-skepticism, & x-risk thinking
- Outlier affinity with academic theory groups, esp. given our size
- Fully decentralized: no offices; team members residing in 20 countries
- Market-leading compensation and benefits for full-timers
  - even by San Francisco Bay Area tech standards
    - (unless you're a machine-learning practitioner)

Protocol Labs **Research**

# (Some of) our ACT-related research interest areas

**Decentralized collaborative editing** of **causal models**

- Categories of **dynamical systems**, causal models, **MDPs**, **POMDPs**
- **Combinators** for the above (e.g. along the lines of AlgebraicJulia's talks)
- Convergent replicated datatypes (**CRDTs**) as categories
- **String-diagram rewriting** (e.g. **double pushouts** of **hypernets**)
- Semi-automated **conflict resolution** of string-diagram rewrites/edits
- **Schema evolution** and Bx applied to modifying generators/equations
- Efficient, *incremental*, semantically guaranteed **implementations** of:
  - **probabilistic programming** inference
  - **value iteration** and **policy iteration**
  - other (PO)MDP **model-checking** algorithms

Protocol Labs **Research**

# (Some of) our ACT-related research interest areas

## Decentralized databases

- Decentralized query planning via rewriting generalized string diagrams
  - using `rewalt`? need to fiber execution steps over diagram of network topology
- Semantic/categorical bridges between **query languages** & **type theory**
- Building a theory around the practice of **hash-linked data**
- Efficient **representations for diffs** that encompass **schema diffs**
- **Incremental query evaluation** with respect to such diffs (cf. CALM theorem); incremental queries as internal functors?
- Generally—**bridges between:**
  - Bx and (enriched) lenses
  - Optics and dependent optics
  - Change Actions
  - CALM theorem
  - CRDTs
  - LVars
  - etc.

Protocol Labs **Research**

# (Some of) our ACT-related research interest areas

**Mechanism design:** strategyproofness and Pareto-efficiency

- Probabilistic social-choice theory with convex algebras ($\mathrm{EM}(\Delta)$)
- **Preference aggregation** schemes as products in certain categories
- Compositional analysis of **sequential collective choices**
- Compositional **credit assignment** within coalitions
  - **Shapley value** as an operad algebra?
- Compositional **bargaining solutions**

Protocol Labs **Research**

# (Some of) our ACT-related research interest areas

## AI existential safety

- Better formalizations of concepts like
  - the orthogonality thesis
  - convergent instrumental goals
  - goal-directedness
- **Eliciting Latent Knowledge** with final coalgebras

Supposing that the machine and the human are working with the same observation space ($O := \mathrm{CameraState}$) and action space ($A := \mathrm{Action}$), then the human's model $H : S_H \to A \to \mathcal{P}(O \times S_H)$ and the machine's model $M : S_M \to A \to \mathcal{P}(O \times S_M)$ are both coalgebras of the endofunctor $F := \lambda X.\, A \to \mathcal{P}(O \times X)$, therefore both have a canonical morphism into the terminal coalgebra of $F$, $X :\cong FX$ (assuming that such an $X$ exists in the ambient category). That is, we can map $S_H \to X$ and $S_M \to X$. Then, if we can define a distance function on $X$ with type $d_X : X \times X \to \mathbb{R}^{\geq 0}$, we can use these maps to define distances between human states and machine states, $d : S_H \times S_M \to \mathbb{R}^{\geq 0}$.

**How can we make use of a distance function?** Basically, we can use the distance function to define a kernel (e.g. $K(x, y) = \exp(-\beta d_X(x, y))$), and then use kernel regression to predict the utility of states in $S_M$ by averaging "nearby" states in $S_H$, and then finally (and crucially) estimating the generalization error so that states from $S_M$ that aren't really near to *anywhere* in $S_H$ get big warning flags (and/or utility penalties for being outside a trust region).

**How to get such a distance function?** One way is to use CMet (the category of complete metric spaces) as the ambient category, and instantiate $\mathcal{P}$ as the Kantorovich monad. Crank-turning yields the formula

$$d_X(s_H, s_M) = \sup_{a:A} \sup_{U:O\times X \to \mathbb{R}} \left| \mathbb{E}_{o,s'_H \sim H(s_H)(a)} U(o, s'_H) - \mathbb{E}_{o,s'_M \sim M(s_M)(a)} U(o, s'_M) \right|$$

Protocol Labs **Research**

# Ways you can work with us

- **Grants**
  - Doctoral candidate fellowships
  - Postdoctoral fellowships
  - Faculty research grants
  - Faculty sabbatical awards
  - See `https://grants.protocol.ai`
- **Part-time** "scoped contributor" roles
  - Anywhere between 50% and 100% time
  - Submit a brief workplan to address a specific problem for 2–6 months
- **Full-time** Research Scientist role
  - If our interests overlap enough that there's no danger of depletion any time soon!
- Email `davidad@protocol.ai` or tweet `@davidad`

Protocol Labs
**Research**

# Ways you can work with us

- **Grants**
  - Doctoral candidate fellowships
  - Postdoctoral fellowships
  - Faculty research grants
  - Faculty sabbatical awards
  - See `https://grants.protocol.ai`
- **Part-time** "scoped contributor" roles
  - Anywhere between 50% and 100% time
  - Submit a brief workplan to address a specific problem for 2–6 months
- **Full-time** Research Scientist role
  - If our interests overlap enough that there's no danger of depletion any time soon!
- Email `davidad@protocol.ai` or tweet `@davidad`

Thank you!

Protocol Labs **Research**