String Diagrams for Layered Explanations

Leo Lobski and Fabio Zanasi

University College London

Applied Category Theory 2022
18 July

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]

---

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]
3. Blame analysis[3]

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]
3. Blame analysis[3]
4. Artificial intelligence[4]

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]
3. Blame analysis[3]
4. Artificial intelligence[4]

Choosing the right level of abstraction of paramount importance in all of the above, e.g.

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]
3. Blame analysis[3]
4. Artificial intelligence[4]

Choosing the right level of abstraction of paramount importance in all of the above, e.g.

1. Functional vs. mechanistic level

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Explanations and levels of abstraction

Motivation

Ability to provide an explanation is increasingly important in various areas of (computer) science:

1. Data integration in systems biology[1]
2. Autonomous decision making[2]
3. Blame analysis[3]
4. Artificial intelligence[4]

Choosing the right level of abstraction of paramount importance in all of the above, e.g.

1. Functional vs. mechanistic level
2. What level of detail is wanted, who the explanation is given to etc.

---

[1] Krivine, 2017
[2] Nashed et al. 2022
[3] Goessler and Le Metayer 2015
[4] The Royal Society 2019

Layers of abstraction

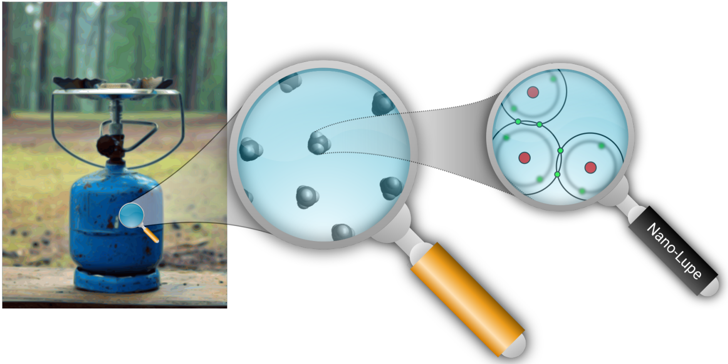


Image source: Openclipart

Structure of explanations

Motivation for formalism

- Want to take layers of abstraction into account

Structure of explanations

Motivation for formalism

- Want to take layers of abstraction into account
- Want to go easily from one context/theory to another

Structure of explanations

Motivation for formalism

- Want to take layers of abstraction into account
- Want to go easily from one context/theory to another
- What is being explained is seen as a *process*

Structure of explanations

Motivation for formalism

- Want to take layers of abstraction into account
- Want to go easily from one context/theory to another
- What is being explained is seen as a *process*
- Processes and their explanations should *compose*

Structure of explanations

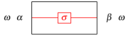
Motivation for formalism

- Want to take layers of abstraction into account
- Want to go easily from one context/theory to another
- What is being explained is seen as a *process*
- Processes and their explanations should *compose*
- Modularity: an explanation of a sub-process should be seen as a partial explanation of the larger process

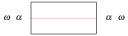
Layered props

$$\frac{\sigma \in \Sigma \setminus \Sigma^i}{\sigma : (\mathtt{ar}(\sigma) \mid \mathtt{coar}(\sigma))}$$

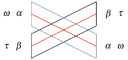
$$\sigma \in \Sigma^i \quad \mathtt{ar}(\sigma) = \omega, \alpha \quad \mathtt{coar}(\sigma) = \omega, \beta$$



$: (\omega, \alpha \mid \omega, \beta)$



$: (\omega, \alpha \mid \omega, \alpha)$

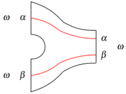


$: (\omega, \alpha; \tau, \beta \mid \tau, \beta; \omega, \alpha)$

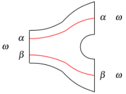
$: (\varepsilon \mid \varepsilon)$

$: (\varepsilon \mid \omega, \varepsilon)$

$: (\omega, \varepsilon \mid \varepsilon)$

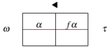


$: (\omega, \alpha; \omega, \beta \mid \omega, \alpha\beta)$



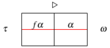
$: (\omega, \alpha\beta \mid \omega, \alpha; \omega, \beta)$

$$f : \omega \to \tau$$



$: (\omega, \alpha \mid \tau, f\alpha)$

$$f : \omega \to \tau$$



$: (\tau, f\alpha \mid \omega, \alpha)$

$$\frac{x : (t \mid s) \quad y : (s \mid u)}{x ; y : (t, u)} \qquad \frac{x : (\omega, \alpha \mid \omega, \gamma) \quad y : (\omega, \beta \mid \omega, \delta)}{x \otimes_\omega y : (\omega, \alpha\beta \mid \omega, \gamma\delta)} \qquad \frac{x : (t \mid s) \quad y : (u \mid w)}{x \otimes y : (t ; u \mid s ; w)}$$

Example: glucose phosphorylation

after Jean Krivine

We wish to explain the chemical reaction rule:

$$glucose + ATP \longrightarrow glucose\text{-}6\text{-}phosphate + ADP + hydrogen\ ion$$

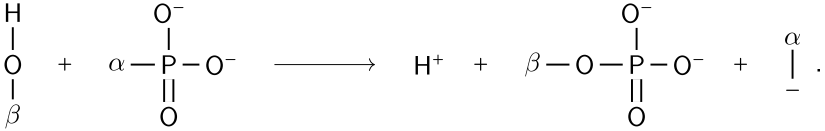
Example: glucose phosphorylation

after Jean Krivine

We wish to explain the chemical reaction rule:

$$\textit{glucose} + \textit{ATP} \longrightarrow \textit{glucose-6-phosphate} + \textit{ADP} + \textit{hydrogen ion}$$

using the interaction between parts of molecules:



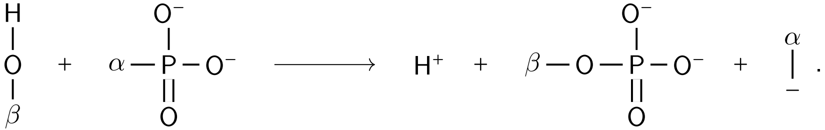
Example: glucose phosphorylation

after Jean Krivine

We wish to explain the chemical reaction rule:

$$glucose + ATP \longrightarrow glucose\text{-}6\text{-}phosphate + ADP + hydrogen\ ion$$

using the interaction between parts of molecules:

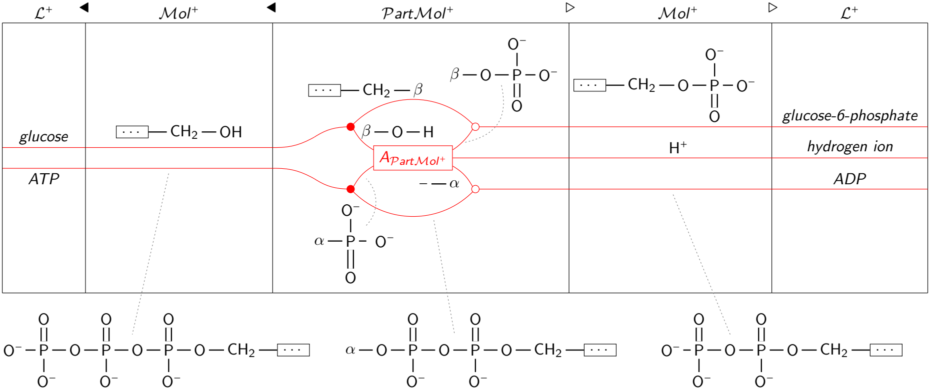


We draw this as an internal box:

in the layered prop.

Example: glucose phosphorylation

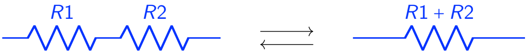
after Jean Krivine



Example: electrical circuits

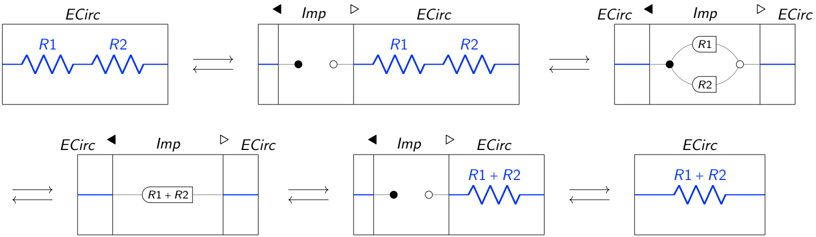
after Boisseau and Sobociński

We wish to explain the rule for the sequential composition of resistors:



Example: electrical circuits

after Boisseau and Sobociński



Example: Calculus of Communicating Systems

after Jean Krivine

We wish to explain the rewrite rule

$$x.0 \parallel (y.0 \parallel \bar{x}.0) \to 0 \parallel (y.0 \parallel 0).$$

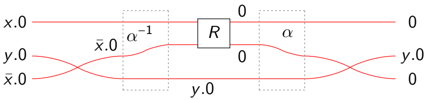
Example: Calculus of Communicating Systems

after Jean Krivine

We wish to explain the rewrite rule

$$x.0 \parallel (y.0 \parallel \bar{x}.0) \to 0 \parallel (y.0 \parallel 0).$$

It has a derivation:



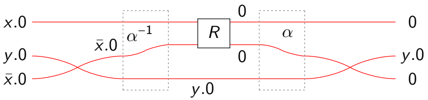
Example: Calculus of Communicating Systems

after Jean Krivine

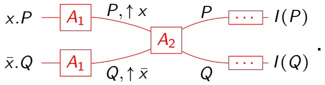
We wish to explain the rewrite rule

$$x.0 \parallel (y.0 \parallel \bar{x}.0) \to 0 \parallel (y.0 \parallel 0).$$

It has a derivation:



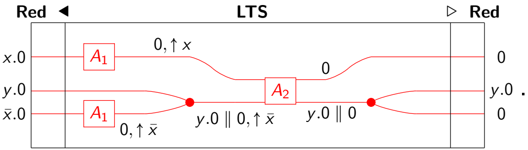
which can be turned into an explanation using the translation:



Example: Calculus of Communicating Systems

after Jean Krivine

There is also a counterfactual explanation:



Definition (Explanation of a 1-cell)

Let $\mathfrak{e}$ and σ be parallel 1-cells in a layered prop. We say that $\mathfrak{e}$ is an *explanation* of σ if

Definition (Explanation of a 1-cell)

Let $\mathfrak{e}$ and σ be parallel 1-cells in a layered prop. We say that $\mathfrak{e}$ is an *explanation* of σ if

1. σ is an internal morphism contained in some category $\omega \in \Omega$,

Definition (Explanation of a 1-cell)

Let $\mathfrak{e}$ and σ be parallel 1-cells in a layered prop. We say that $\mathfrak{e}$ is an *explanation* of σ if

1. σ is an internal morphism contained in some category $\omega \in \Omega$,
2. every internal non-identity morphism of $\mathfrak{e}$ is contained in some category ω' such that $\omega' < \omega$ in the partial order of Ω,

Definition (Explanation of a 1-cell)

Let $\mathfrak{e}$ and σ be parallel 1-cells in a layered prop. We say that $\mathfrak{e}$ is an *explanation* of σ if

1. σ is an internal morphism contained in some category $\omega \in \Omega$,
2. every internal non-identity morphism of $\mathfrak{e}$ is contained in some category ω' such that $\omega' < \omega$ in the partial order of Ω,
3. there is either a 2-cell $\mathfrak{e} \to \sigma$ or a 2-cell $\sigma \to \mathfrak{e}$.

Definition (Explanation of a 2-cell)

Let η and μ be parallel 2-cells in a layered prop. We say that η is an *explanation* of μ if

Explanations

Formal definitions

Definition (Explanation of a 2-cell)

Let η and μ be parallel 2-cells in a layered prop. We say that η is an *explanation* of μ if

1. μ is generated by an equality of morphisms in some category $\omega \in \Omega$,

Explanations

Formal definitions

Definition (Explanation of a 2-cell)

Let η and μ be parallel 2-cells in a layered prop. We say that η is an *explanation* of μ if

1. μ is generated by an equality of morphisms in some category $\omega \in \Omega$,
2. η can be constructed using the generating 2-cells of a layered prop and the 2-cells that come from an equality of morphisms in those categories ω' for which $\omega' < \omega$ in the partial order of Ω.

Explanations

Formal definitions

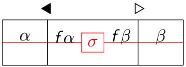
Definition (Explanation of a 2-cell)

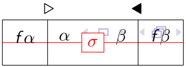
Let η and μ be parallel 2-cells in a layered prop. We say that η is an *explanation* of μ if

1. μ is generated by an equality of morphisms in some category $\omega \in \Omega$,
2. η can be constructed using the generating 2-cells of a layered prop and the 2-cells that come from an equality of morphisms in those categories ω' for which $\omega' < \omega$ in the partial order of Ω.

Definition (Window, cowindow)

A *window* is a morphism in a layered prop of the form on the left below. Dually, a *cowindow* is a morphism in a layered prop of the form on the right below.



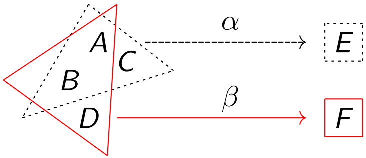


Future directions

- Data integration and hypothesis generation

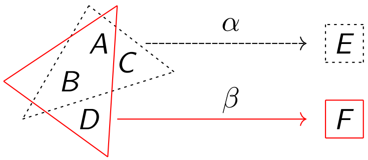
Future directions

- Data integration and hypothesis generation
- Unsharp compartments in biology



Future directions

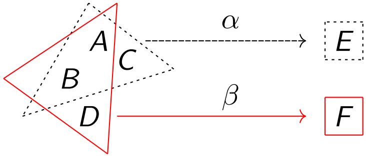
- Data integration and hypothesis generation
- Unsharp compartments in biology



- Formalising scientific explanations

Future directions

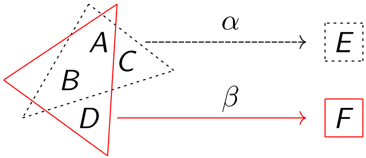
- Data integration and hypothesis generation
- Unsharp compartments in biology



- Formalising scientific explanations
- Counterfactual reasoning

Future directions

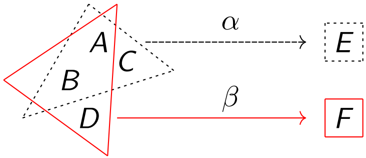
- Data integration and hypothesis generation
- Unsharp compartments in biology



- Formalising scientific explanations
- Counterfactual reasoning
- Layered props as a Grothendieck construction

Future directions

- Data integration and hypothesis generation
- Unsharp compartments in biology



- Formalising scientific explanations
- Counterfactual reasoning
- Layered props as a Grothendieck construction
- Other ideas welcome!

References

- Jean Krivine. *Systems Biology*. ACM SIGLOG News 4(3) 2017.
- Jean Krivine. *Physical systems, composite explanations and diagrams*. SYCO 5, 2019.
- Samer B. Nashed, Saaduddin Mahmud, Claudia V. Goldman, Shlomo Zilberstein. *A Unifying Framework for Causal Explanation of Sequential Decision Making*. arXiv:2205.15462v1 2022.
- Gregor Goessler and Daniel Le Metayer. *A general framework for blaming in component-based systems*. Science of computer programming 113(Part 3) 2015.
- The Royal Society. *Explainable AI: the Basics - Policy Briefing*. 2019.
- Guillaume Boisseau and Paweł Sobociński. *String Diagrammatic Electrical Circuit Theory*. arXiv:2106.07763 2021.

Thank you for your attention!