



**HAL**  
open science

# A semiotic-based framework to assess mental models of XAI systems

Clément Arlotti, Nicolas Heulot

► **To cite this version:**

Clément Arlotti, Nicolas Heulot. A semiotic-based framework to assess mental models of XAI systems. HyCHA 2024 Hybridation Connaissances, Humain et Apprentissage Statistique, Mar 2024, Gif-sur-Yvette, France. hal-04520749

**HAL Id: hal-04520749**

**<https://hal.science/hal-04520749v1>**

Submitted on 25 Mar 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A semiotic-based framework to assess mental models of XAI systems.

Clément ARLOTTI, Nicolas HEULOT

IRT SystemX, site Nano-INNOV, 2 Bd Thomas Gobert, 91120 Palaiseau

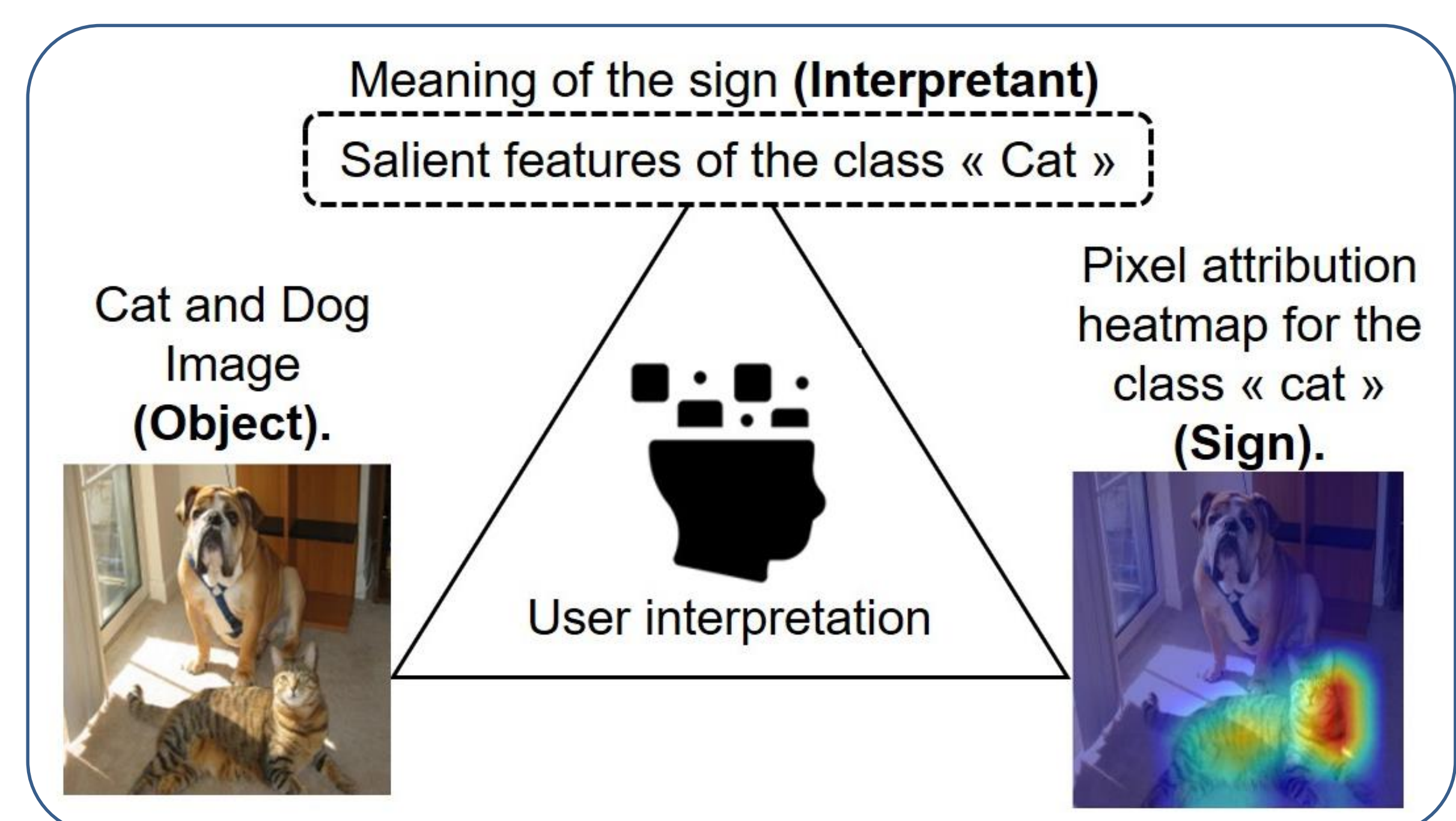
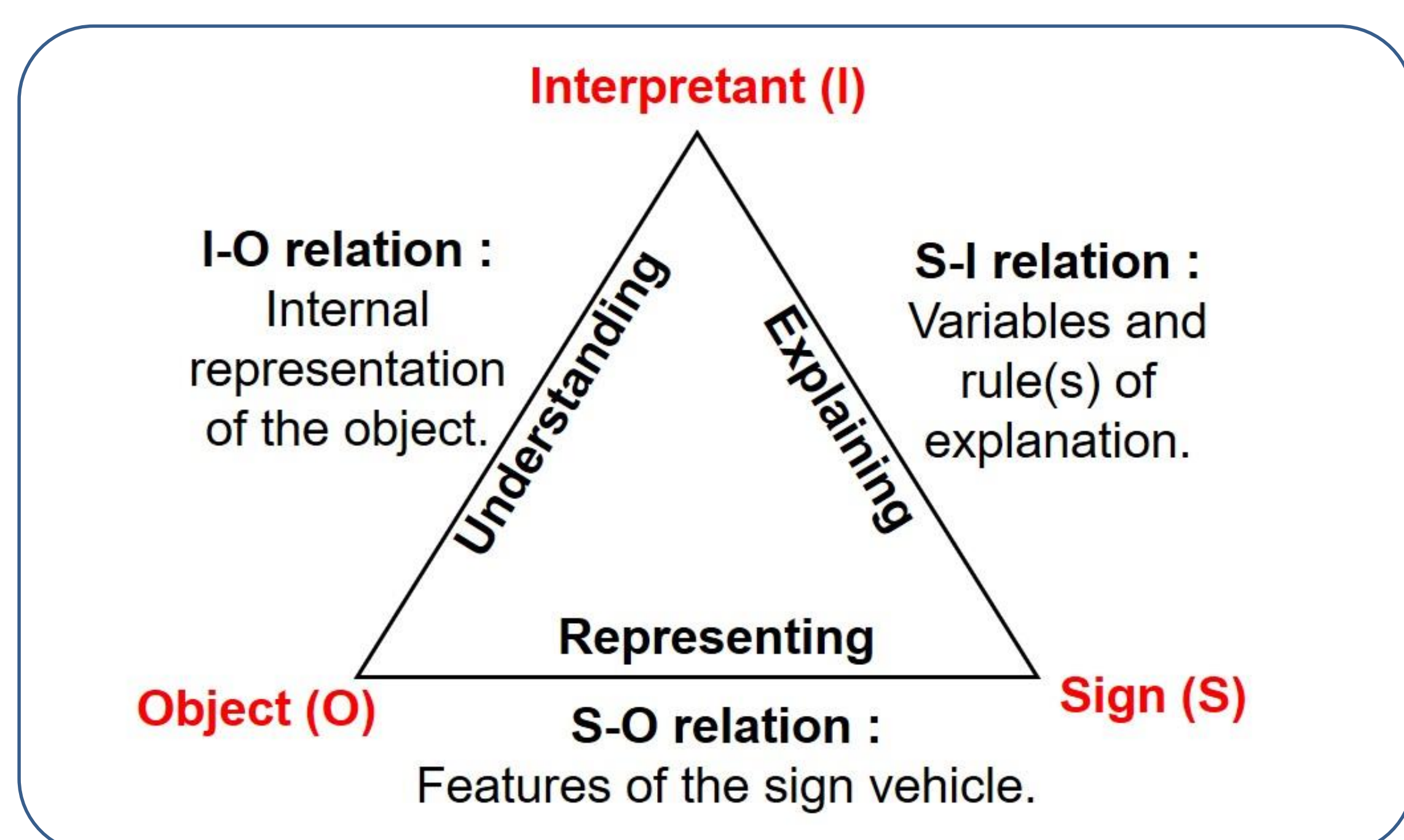
### 1. Introduction

- Most of current XAI systems are algorithm-centric and hardly fit the way of thinking of their end-users.
- How can we describe the way human stakeholders picture such systems ? How to characterize their mental models ?
- Semiotic concepts [1], [2] can be used to describe mental models and connect interdisciplinary elements for XAI.

### 2. What are mental models used for ?

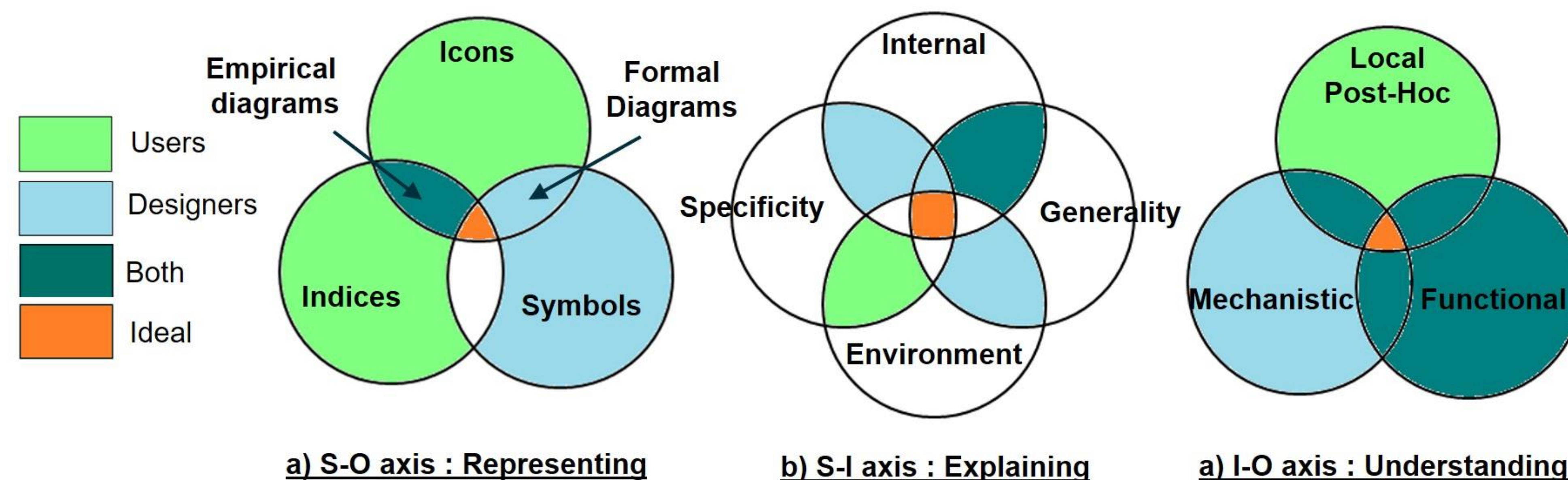
- **Representing** the system in its environment.
- **Understanding** the system's functioning.
- **Explaining** the system's behavior [3], [4].

### 3. Can we use a semiotic approach to describe mental models ?



### 4. How to assess mental models using semiotics ?

- **Goal:** characterize stakeholders mental models components and assess their (mis-)alignment.
- **Approach:** connection of interdisciplinary concepts [5], [6], [7] to embody formal aspects of the semiotic theory.
- **Use case:** computer-vision anomaly detection system to monitor the quality of control points in assembly production line.
- **Method:** group interviews with **Users** and **Designers** of the system.
- **Results overview:**
  - **Users:** mostly use empirical diagrams, mechanism-agnostic context-centric explanations, post-hoc understanding.
  - **Designers:** mostly use formal diagrams context-agnostic mechanism-centric explanations, mechanistic understanding.



### 5. Conclusion

- Semiotic-based, user-centered framework to assess (mis-)aligned mental models components.
- Ability to delineate consistent stakeholder profile tendencies based on semiotic categories.
- Highlight the potential of existing work in semiotics to connect interdisciplinary concepts in XAI.

### Bibliography

- [1] Peirce C.S., Hoopes J. *Peirce on signs: writings on semiotic*, University of North Carolina Press (2006).  
 [2] Nöth W., *The semiotics of models*. Sign Systems Studies (2018)  
 [3] Rutjes et.al., *Considerations on explainable AI and users' mental models*. Glasgow Association for Computing Machinery (2019)  
 [4] Greca et.al., *Mental models, conceptual models, and modelling*. International Journal of Science Education. (2000)  
 [5] Paéz A., *The Pragmatic Turn in Explainable Artificial Intelligence*. Minds and Machines. (2019)  
 [6] Erasmus et.al., *What is Interpretability?* Philosophy & Technology. (2021)  
 [7] Buijsman S. *Defining explanation and explanatory depth in XAI*. Minds and Machines. (2022)