NC STATE UNIVERSITY

TILBURG UNIVERSITY

THE UNIVERSITY OF UTAH

The Visual Narrative Engine

A Computational Model of the Visual Narrative Parallel Architecture

Chris Martens, North Carolina State University
Rogelio E. Cardona-Rivera, University of Utah
Neil Cohn, Tilburg University

rogelio@cs.utah.edu
http://rogel.io
@recardona

How do people understand visual narrative?


CHOOM!

We don’t know the process that underlies visual story understanding.

We don't know the process that underlies ~~visual~~ story understanding.

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist* …they are rational hypotheses, and that's okay

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist* …they are rat[X]nal hypotheses, and that's okay

human unconstrained

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*

…they are rat[X]nal hypotheses, and that's okay

human unconstrained

- Knowledge-lean Story Understanding: Document Analysis
- Knowledge-rich Story Generation: Narrative-theoretic Heuristic Search Planning
- Neurosymbolic Understanding+Generation: Benchmarked Commonsense Reasoning

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*
  - Remarkable progress on story understanding

Mueller, Erik T. "Computational models of narrative." Sprache und Datenverarbeitung: International Journal of Language Processing 37.1-2 (2013): 11-39.

Knowledge-rich Story Understanding

By the 1990s, we knew…

Wilensky *et al.* — people predict which goals and subsequent plans explain observed actions by characters

Norvig — scripts are important for generating knowledge-based inferences

Mueller — spatiotemporal reasoning constrains story inferences

Lehnert — means-ends (causal) and hierarchical (purposive) reasoning contribute the most to a person's memory of a story

Black and Bower — hierarchical problem solving is key for inferencing and understanding

Winston — it is possible to combine these in a principled manner

Knowledge-rich Story Understanding

By the

Wilen

plain

d

s

soning

Black ar

d

Winston— it is possible to combine these in a principled manner

The Psychology Survey says…

They were all right about the concepts!

But not right about the procedures.

Knowledge-rich Story Understanding

By the

Wilen

T

T

B

Black ar

Wi

n

ing

Fun fact:
Neurosymbolic approaches are discovering the same

- Events
- Goals
- Characters
- Scripts

Knowledge-rich Story Understanding

By the

Wilen

n

Black ar

Wi

Another fun fact:
The 80s knew what we had to do

Although the HST theory suggests the form of the products of comprehension in the reader's memory, it is seriously deficient in not spelling out the moment-by-moment *process* by which the reader arrives at those representational products. This deficit is the primary focus for the theoretical work in the future. Although we have no process theory at present, we will indicate some of the considerations and issues that must be resolved in arriving at a process model for story comprehension.

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*
  - Remarkable progress on story understanding

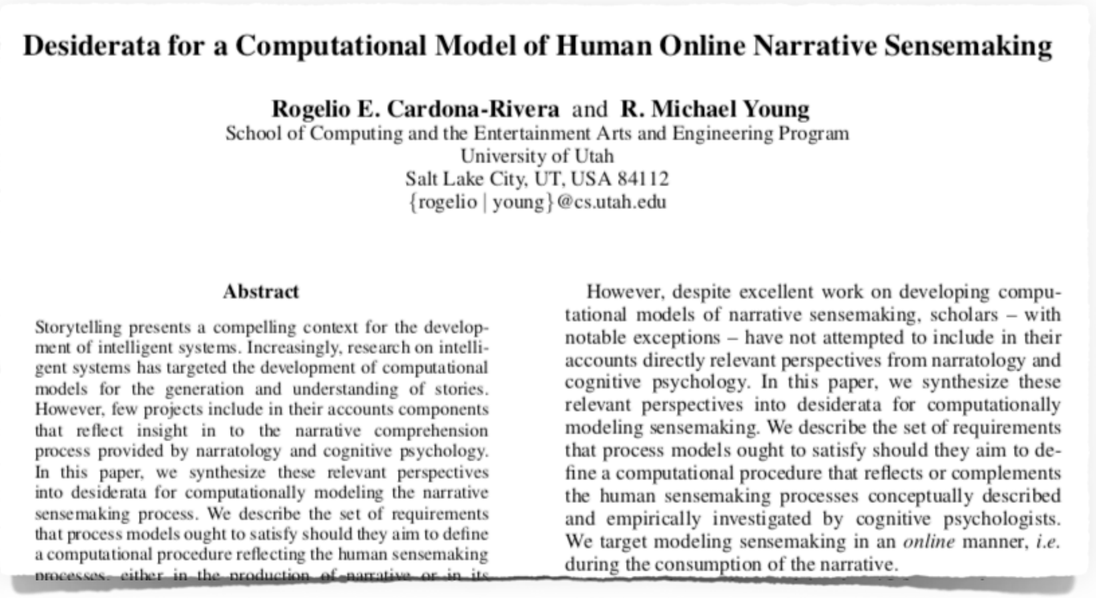
Mueller, Erik T. "Computational models of narrative." Sprache und Datenverarbeitung: International Journal of Language Processing 37.1-2 (2013): 11-39.

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts

We don't know the process that underlies ~~visual~~ story understanding.

- AI is broadly *functionalist*
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
  - Too abstract



Desiderata for a Computational Model of Human Online Narrative Sensemaking

Rogelio E. Cardona-Rivera and **R. Michael Young**
School of Computing and the Entertainment Arts and Engineering Program
University of Utah
Salt Lake City, UT, USA 84112
{rogelio | young}@cs.utah.edu

Abstract

Storytelling presents a compelling context for the development of intelligent systems. Increasingly, research on intelligent systems has targeted the development of computational models for the generation and understanding of stories. However, few projects include in their accounts components that reflect insight in to the narrative comprehension process provided by narratology and cognitive psychology. In this paper, we synthesize these relevant perspectives into desiderata for computationally modeling the narrative sensemaking process. We describe the set of requirements that process models ought to satisfy should they aim to define a computational procedure reflecting the human sensemaking processes, either in the production of narrative or in its

However, despite excellent work on developing computational models of narrative sensemaking, scholars – with notable exceptions – have not attempted to include in their accounts directly relevant perspectives from narratology and cognitive psychology. In this paper, we synthesize these relevant perspectives into desiderata for computationally modeling sensemaking. We describe the set of requirements that process models ought to satisfy should they aim to define a computational procedure that reflects or complements the human sensemaking processes conceptually described and empirically investigated by cognitive psychologists. We target modeling sensemaking in an *online* manner, *i.e.* during the consumption of the narrative.

Rogelio E. Cardona-Rivera and R. Michael Young; Desiderata for a Computational Model of Human Online Narrative Sensemaking. In the *Working Notes of the 2019 AAAI Spring Symposium on Story-enabled Intelligence*, Stanford, CA, USA, 2019.

We don't know the process that underlies ~~visual~~ story understanding.

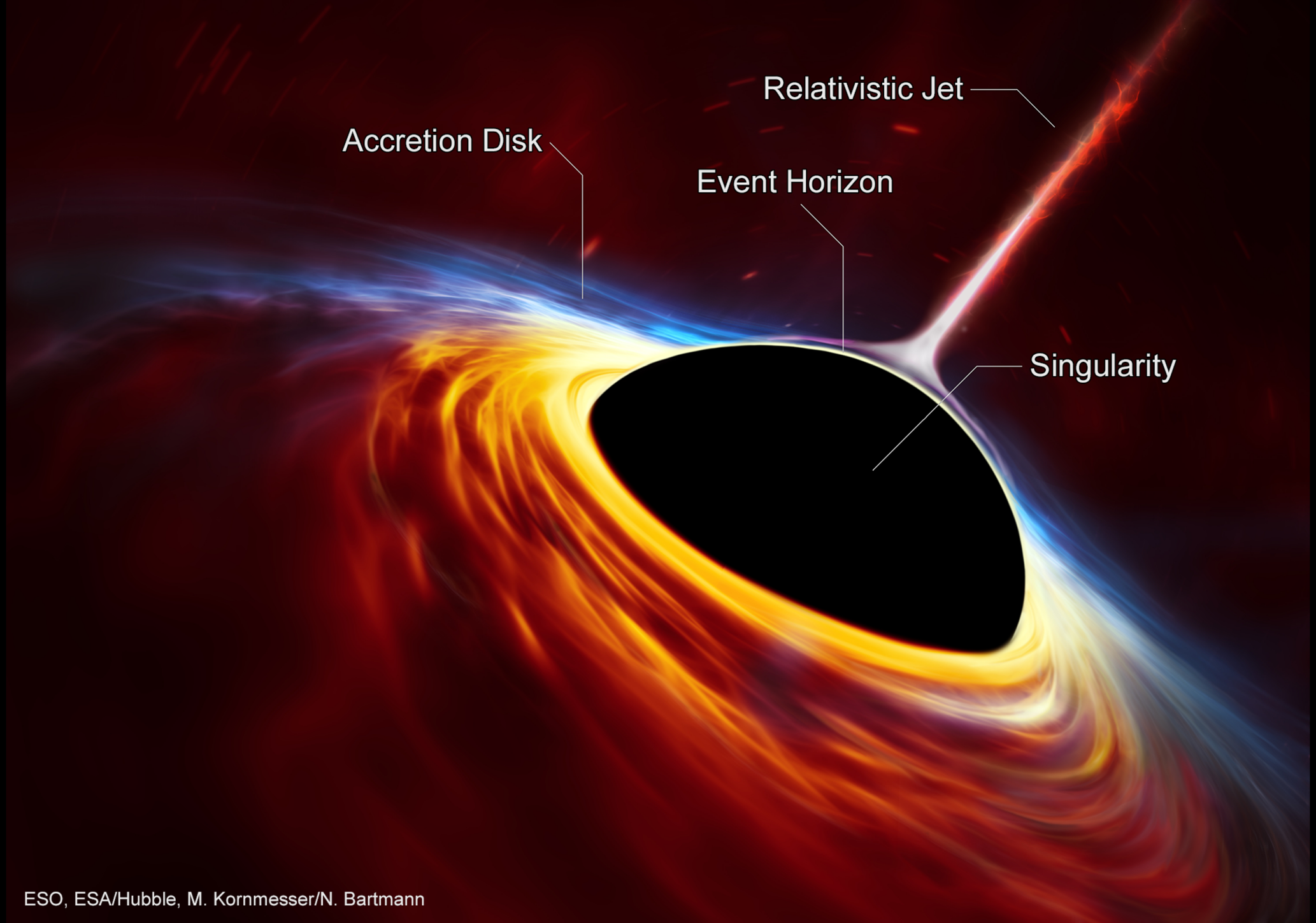
- AI is broadly *functionalist*
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
  - Too abstract
- Soapbox: Must bring these together
  - Biologically-plausible structural models

Lieto, A., & Radicioni, D. P. (2016). From human to artificial cognition and back: New perspectives on cognitively inspired AI systems. Cognitive Systems Research, 39, 1 – 3.

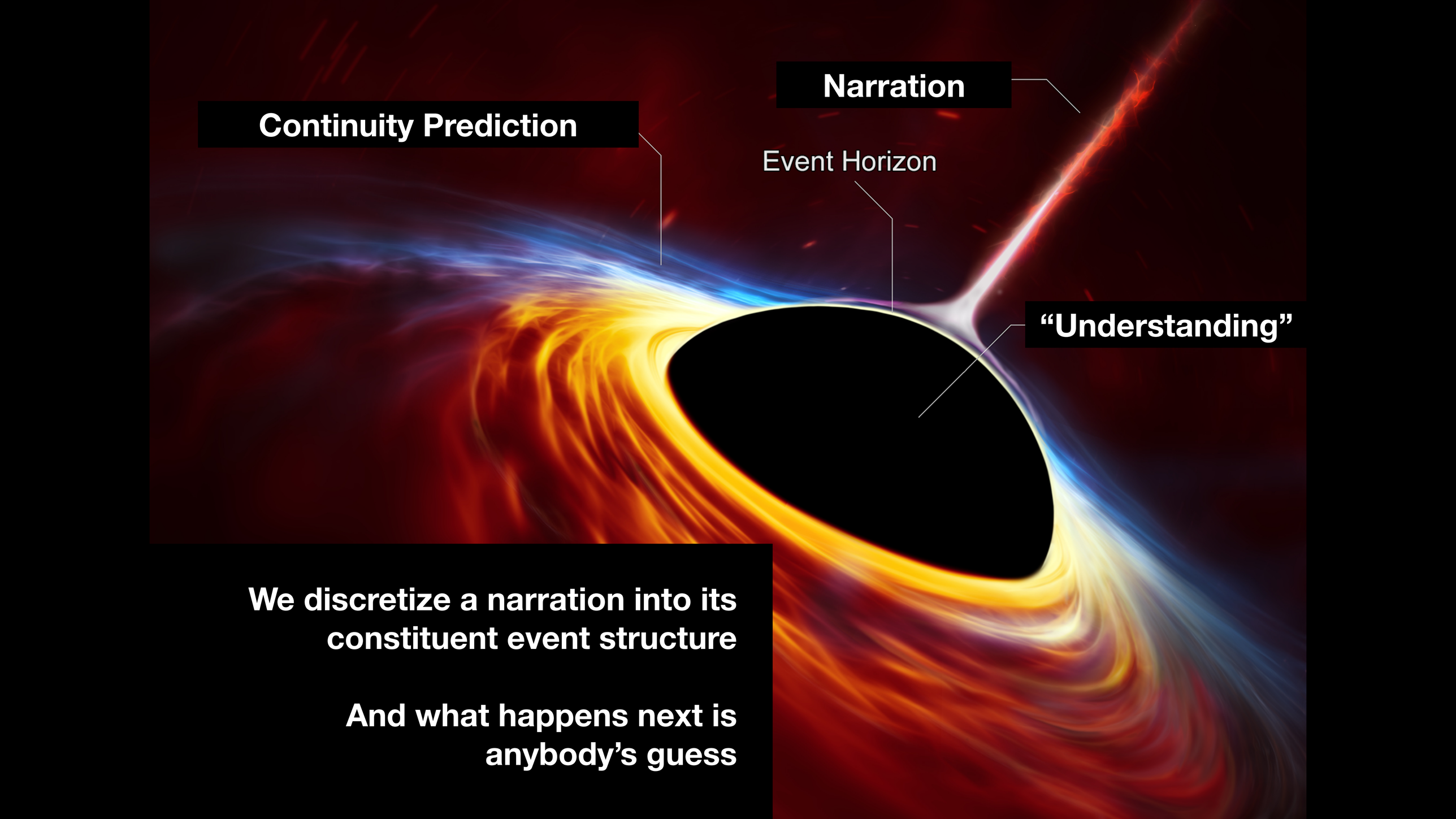
This paper is an existence proof: we can describe the (visual) story understanding procedures mechanically

This paper is an existence proof: we can describe the (visual) story understanding procedures mechanically

We may need to rely on simulation



ESO, ESA/Hubble, M. Kornmesser/N. Bartmann

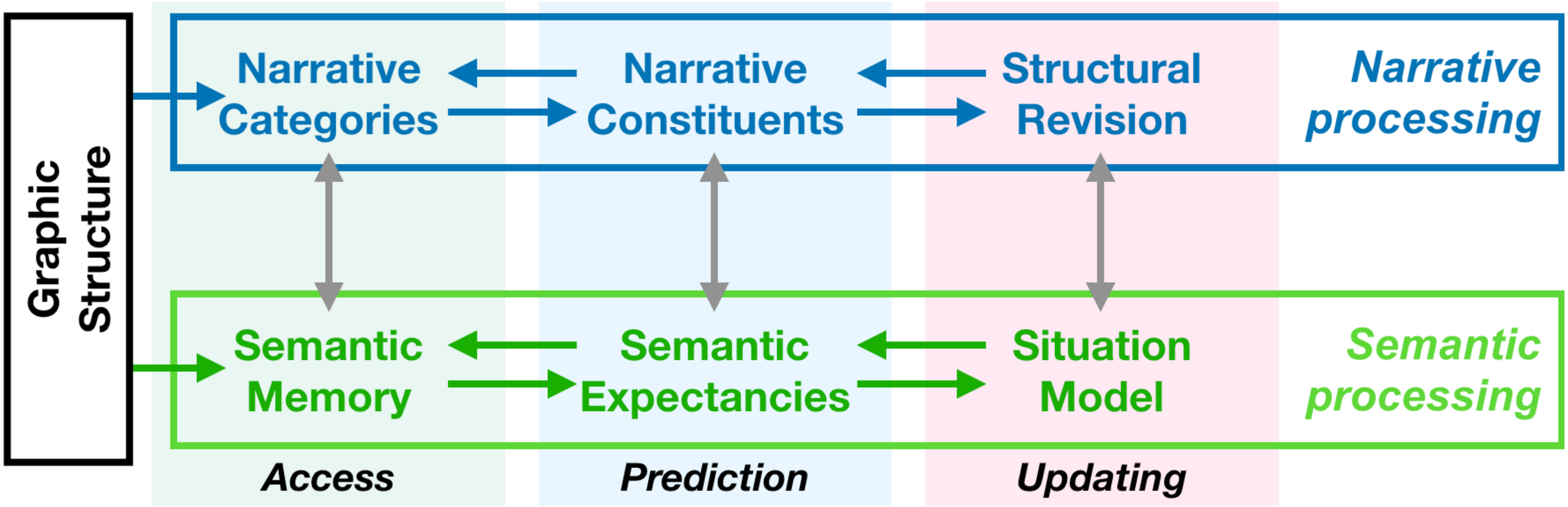


We discretize a narration into its constituent event structure

And what happens next is anybody’s guess

Parallel **I**nterfacing **N**arrative **S**emantics

Dual (Syntax/Semantics) Process Reasoning

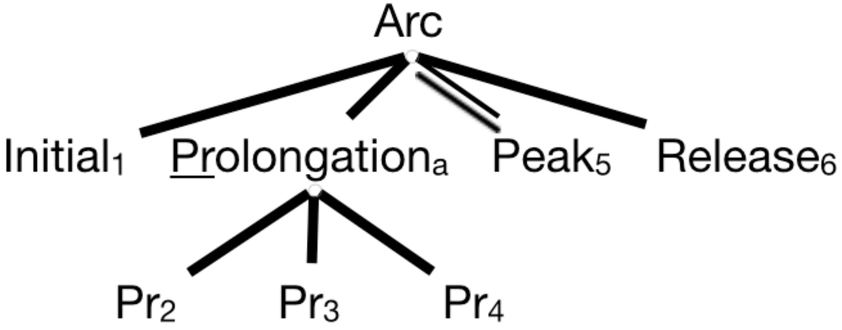


Visual **N**arrative **G**rammar

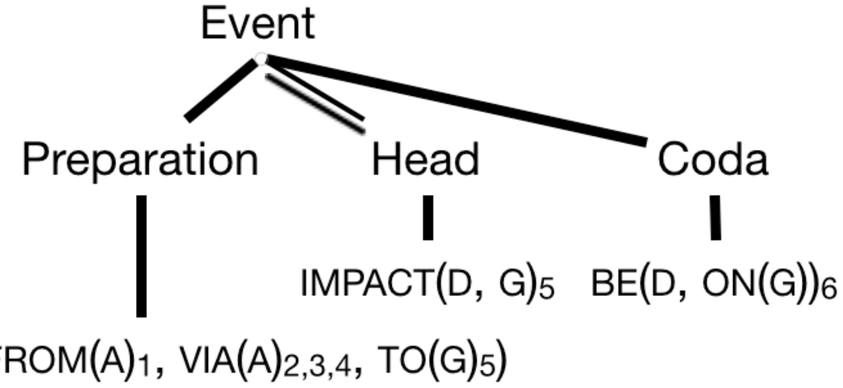
Graphic Structure

1 2 3 4 5 6

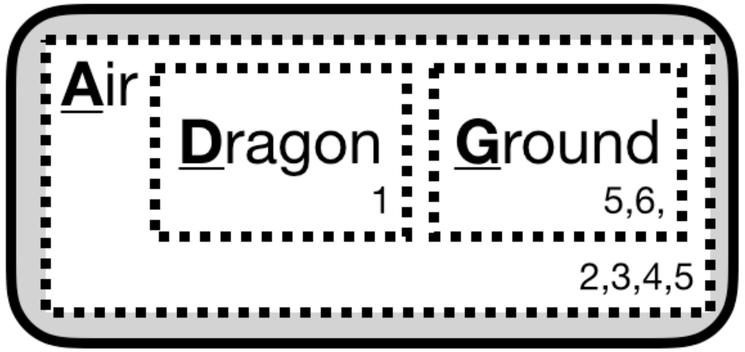
Narrative Structure



Event Structure



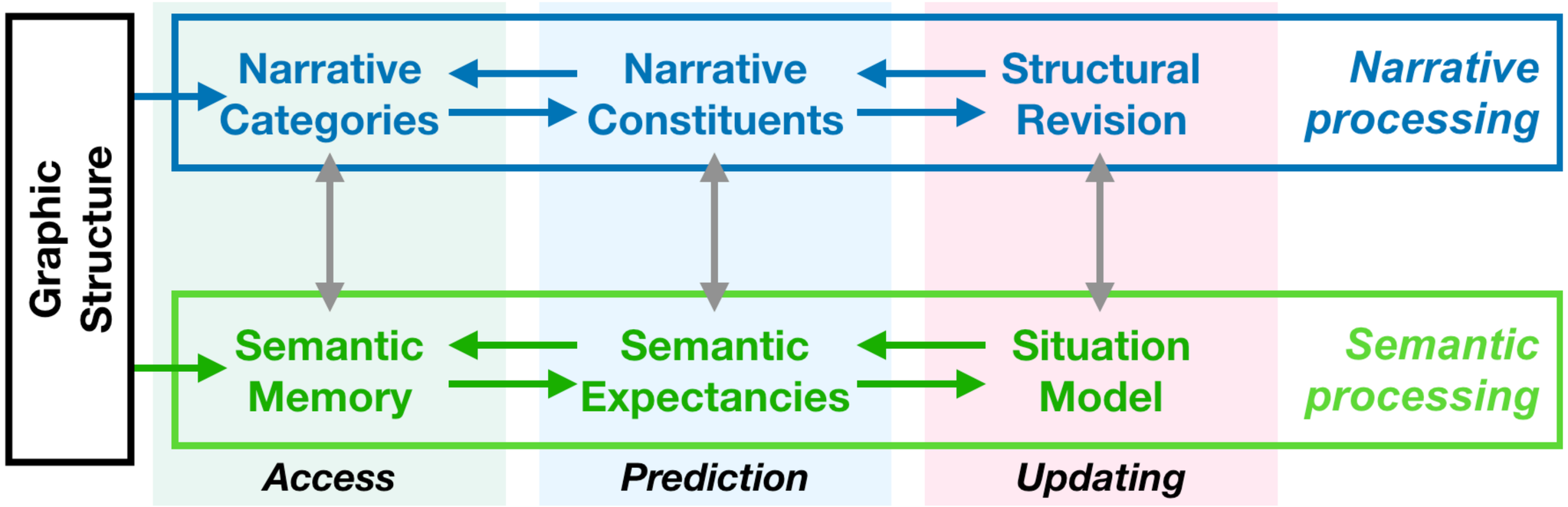
Spatial/Referential Structure



The Visual Narrative Engine

Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

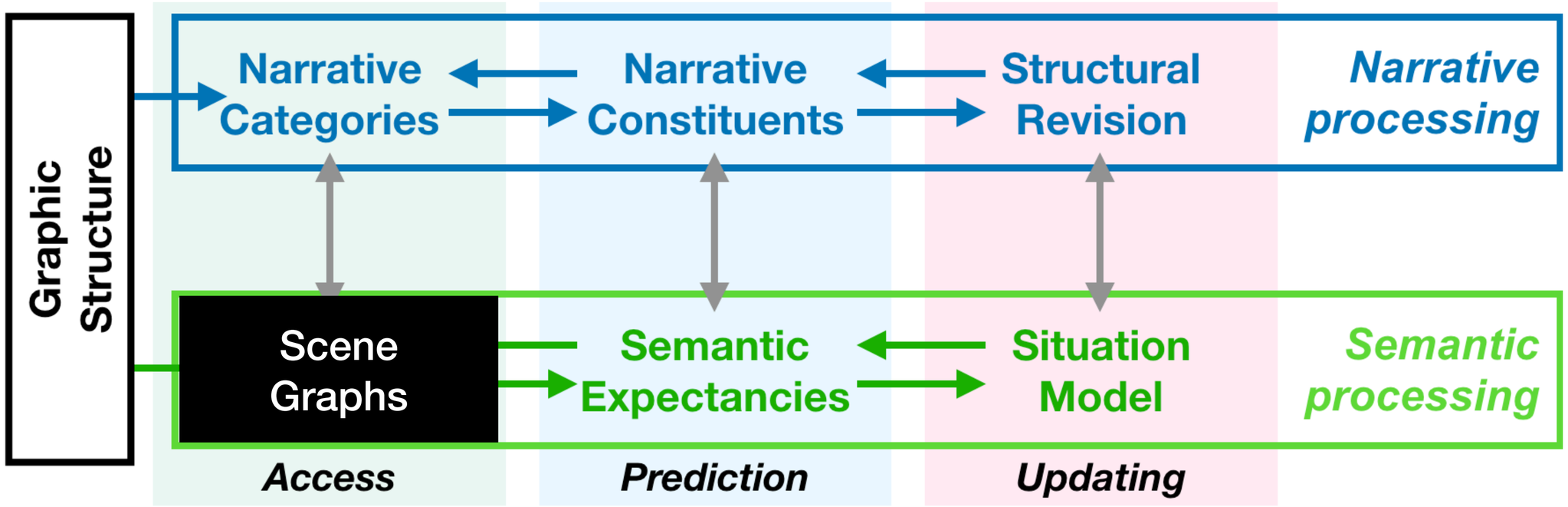
- Research Question: Can we describe procedures to match posited interfaces?



The Visual Narrative Engine

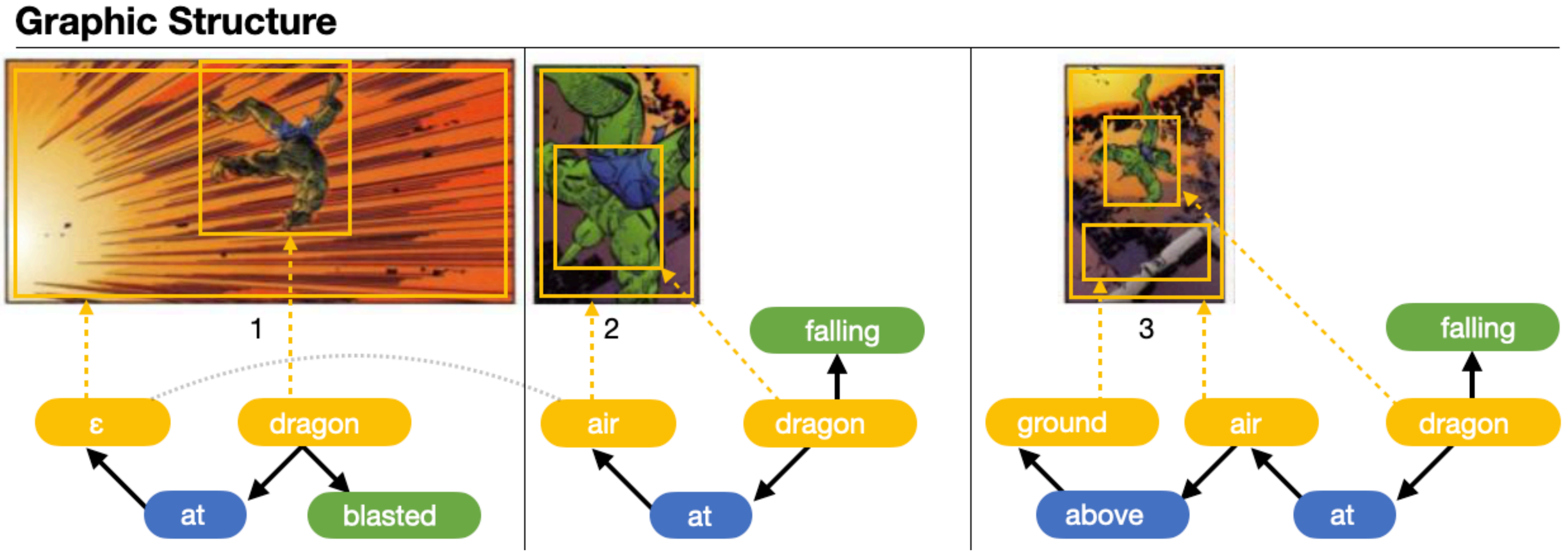
Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

- Research Question: Can we describe procedures to match posited interfaces?



Scene Graphs

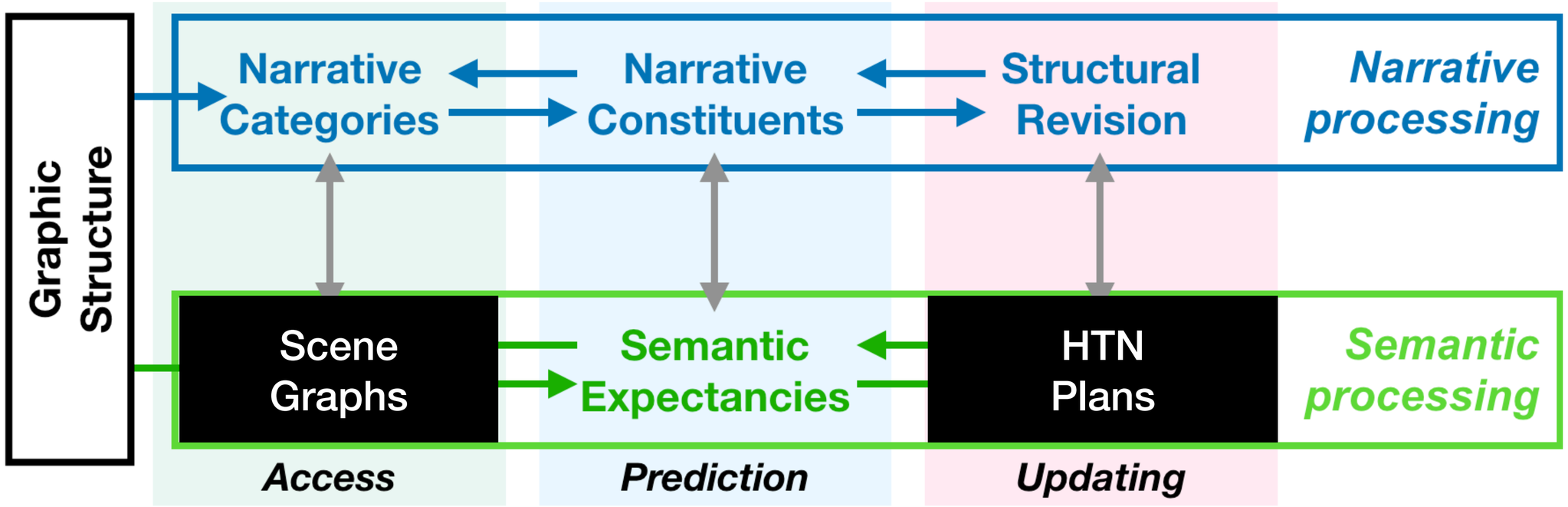
A Representation from Computer Vision



The Visual Narrative Engine

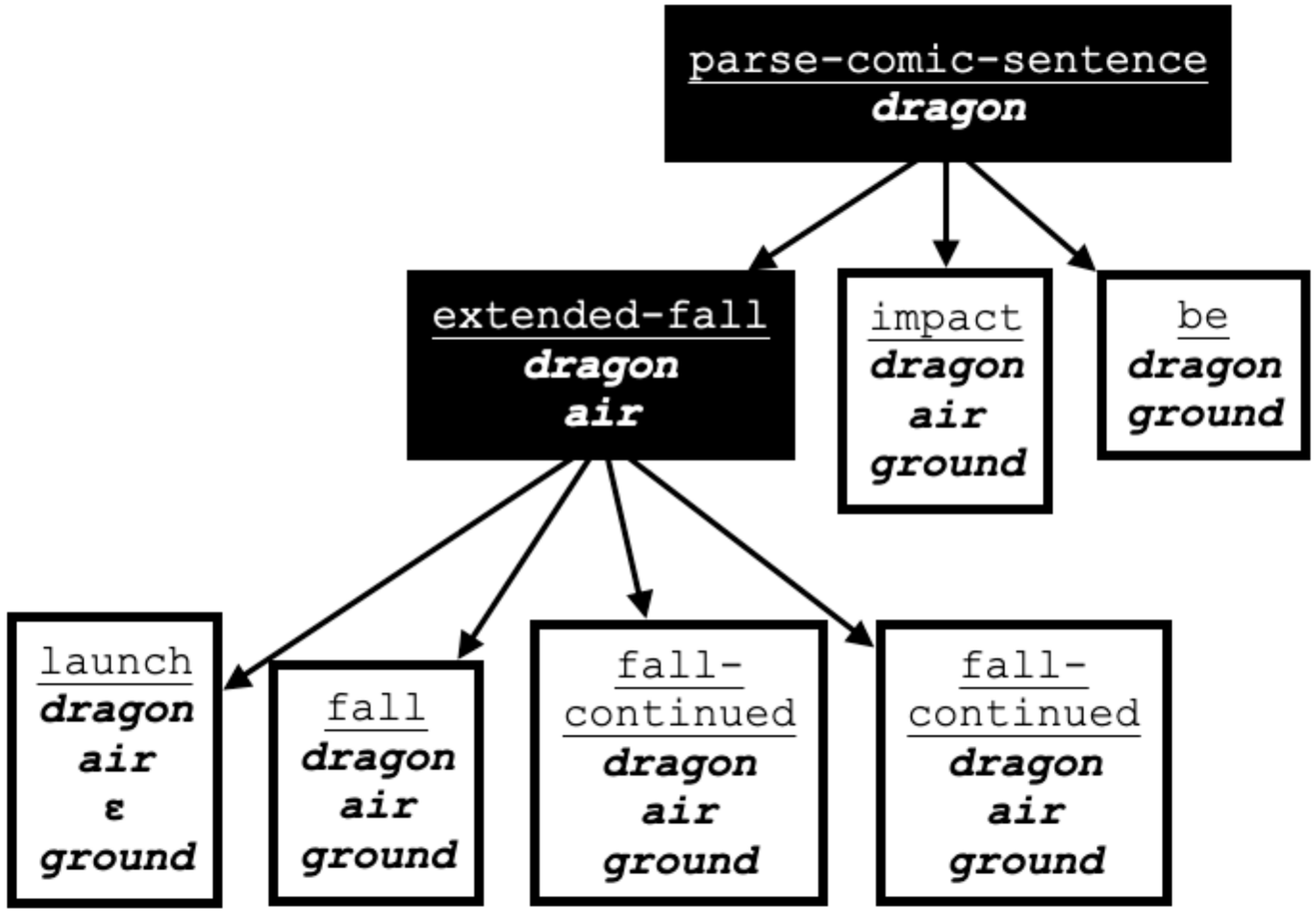
Model of the combined VNG+PINS = Visual Narrative Parallel Arch.

- Research Question: Can we describe procedures to match posited interfaces?

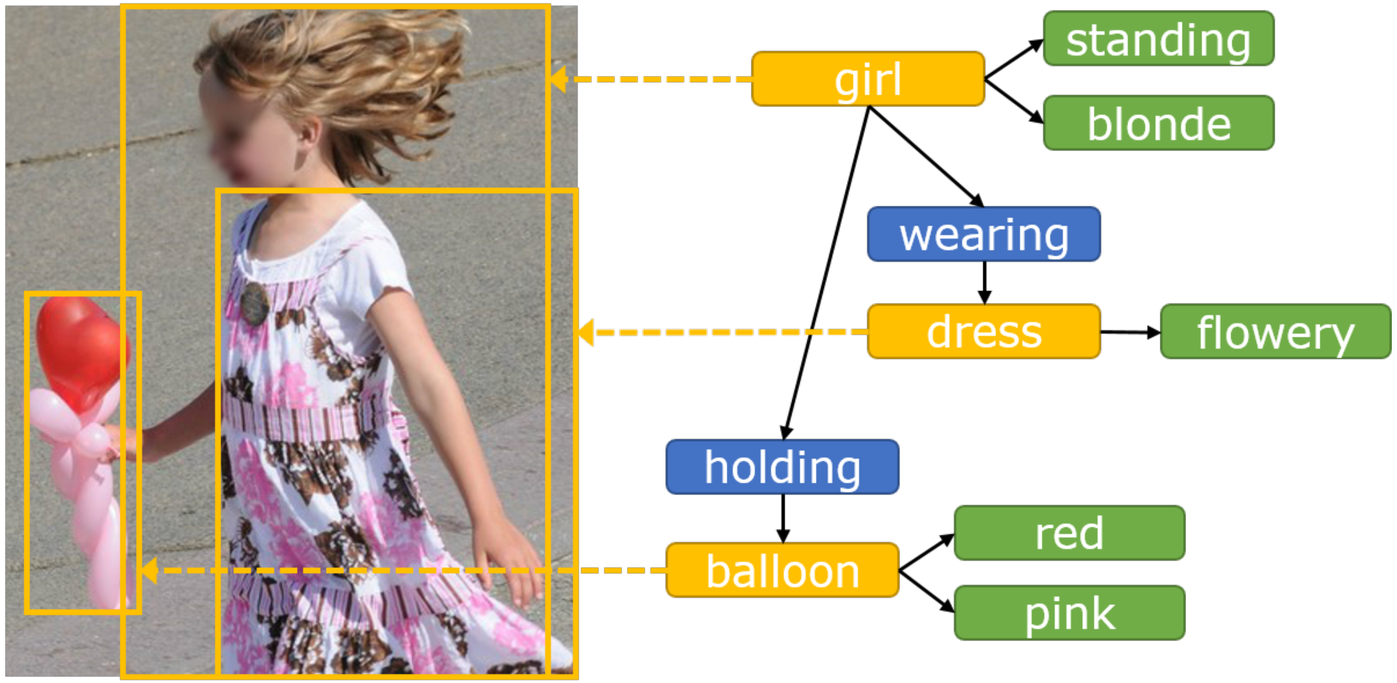


Hierarchical Task Networks

A Representation from Automated Planning



Representations are Compatible!



Binary Literals

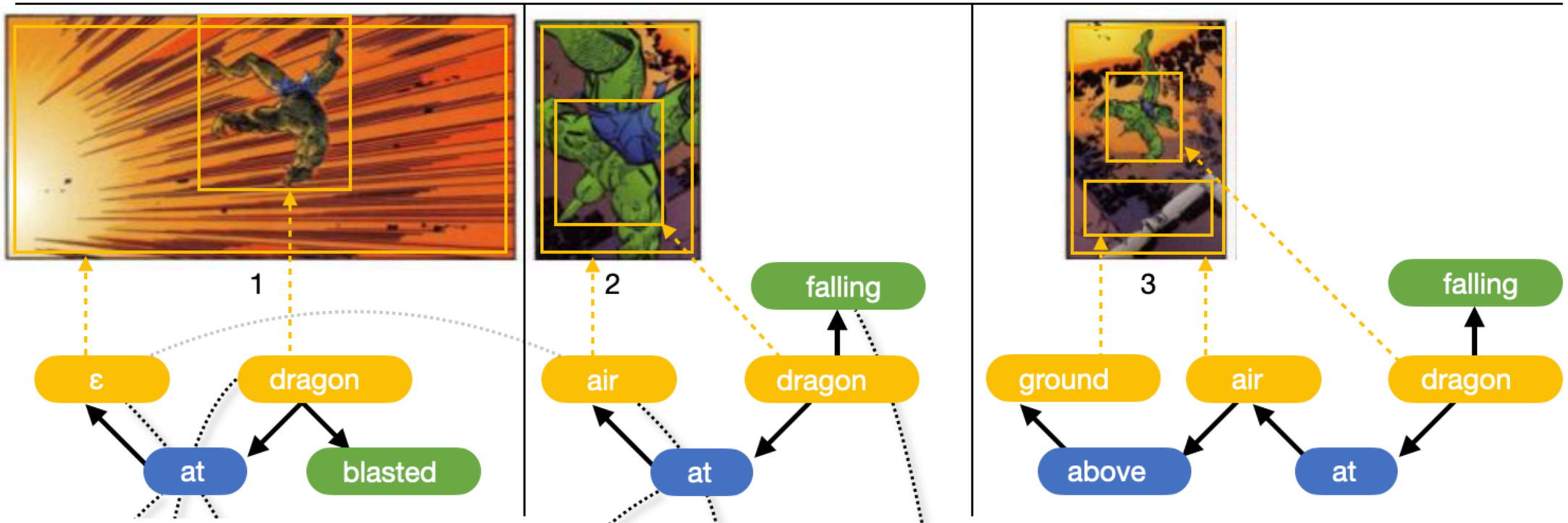
```
wearing(?girl, ?dress)
holding(?girl, ?balloon)
```

Unary Literals

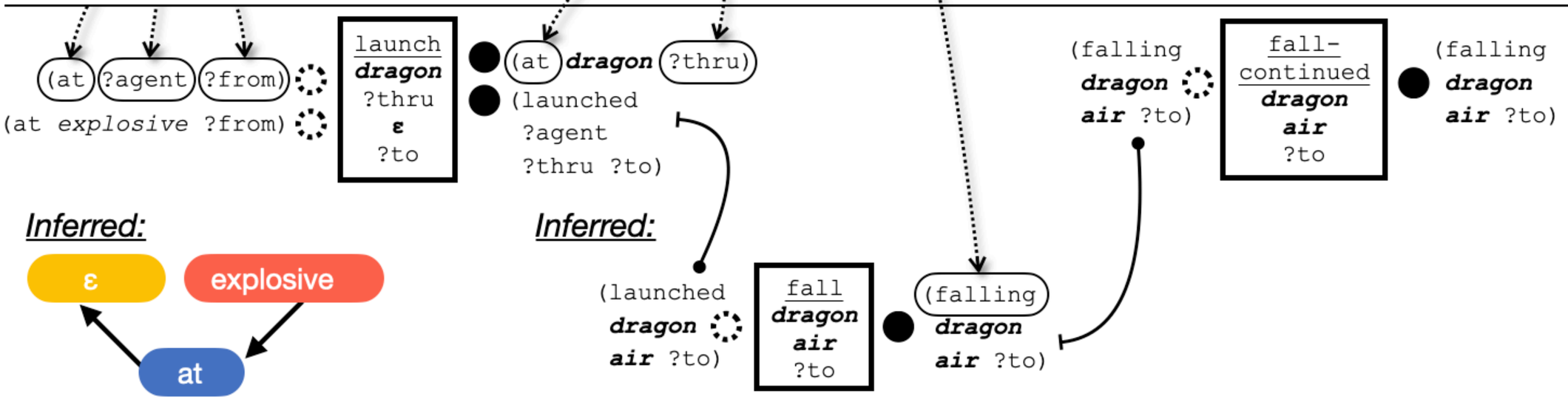
```
standing(?girl)
blonde(?girl)
flowery(?dress)
red(?balloon)
pink(?balloon)
```

Rogelio E. Cardona-Rivera and Boyang Li; PlotShot: Generating Discourse-constrained Stories around Photos. In Proceedings of the 12th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-16), pages 2-8, Burlingame, CA, USA, 2016.

Graphic Structure

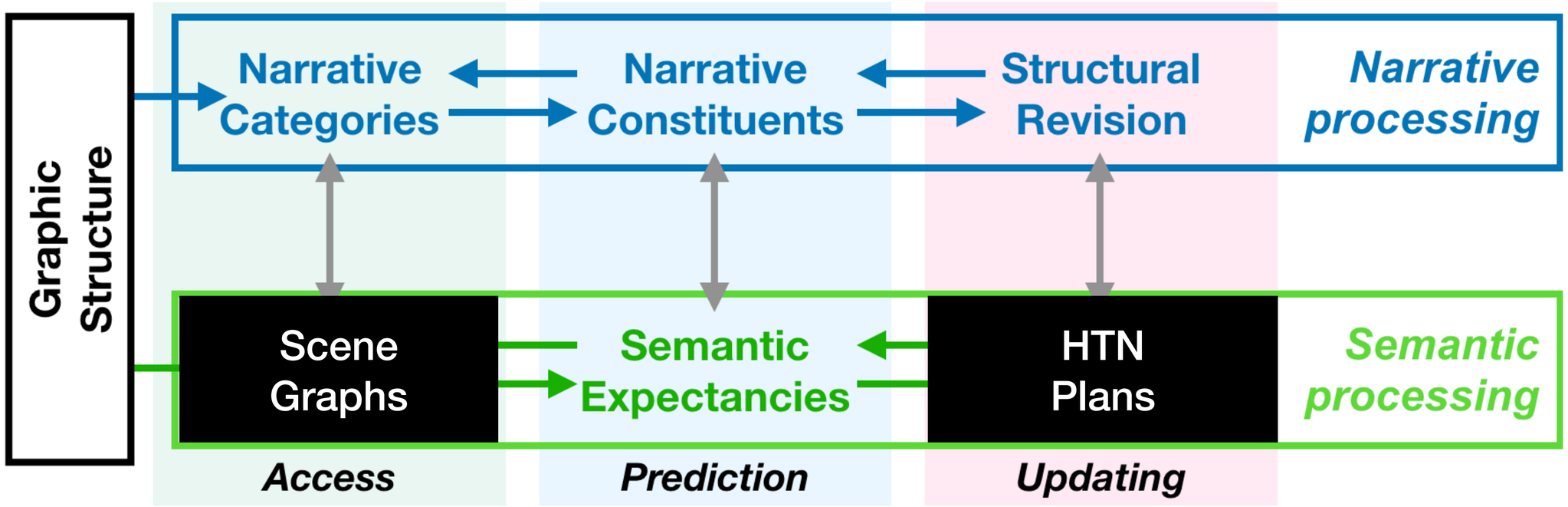


Event Structure

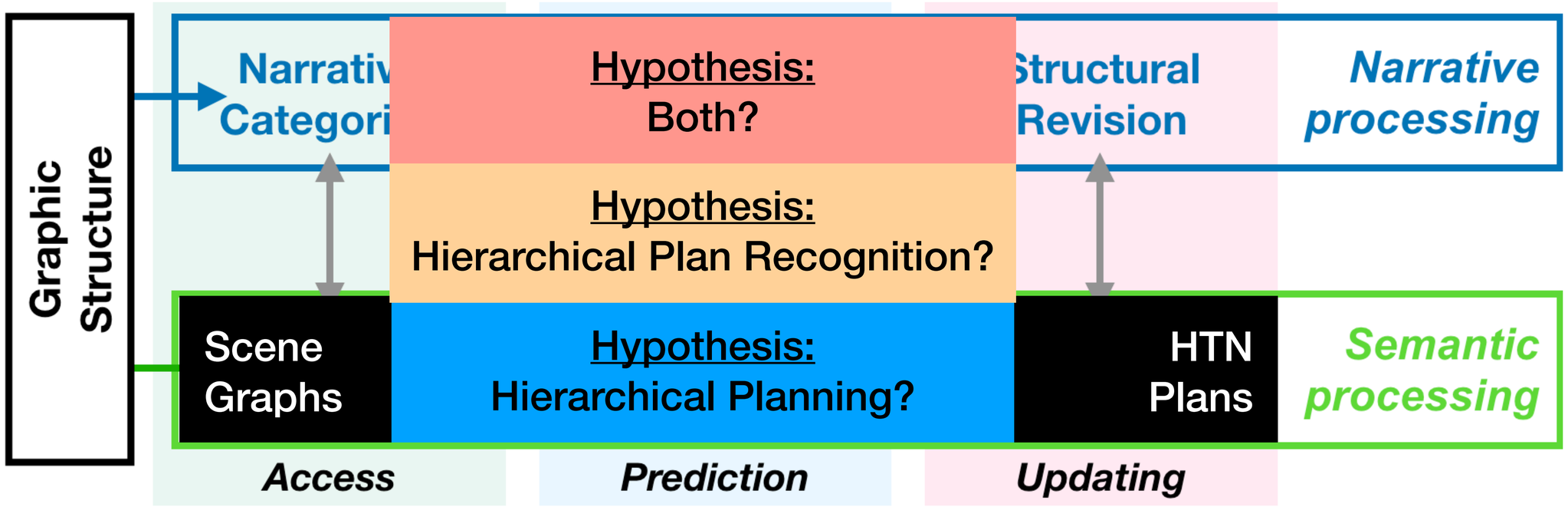


Open question: What's the bridge?

- Research Question: Can we describe procedures to match posited interfaces?



Open question: What's the bridge?



We don’t know the process that underlies visual story understanding.

We don't know the process that underlies visual story understanding.

- AI is broadly *functionalist*
  - Remarkable progress on story understanding
- Story psychology has offered process-level accounts
  - Too abstract
- Soapbox: Must bring these together
  - Biologically-plausible structural models

rogelio@cs.utah.edu

http://rogel.io

@recardona

RECAP