

Identifying Multimodal Errors Through Explanations

Dr. Leilani H. Gilpin
Research Scientist, Sony AI
December 8, 2020

Brief Introduction

- Joined Sony AI in September 2020.
- Education
 - B.S. in Computer Science, B.S. in Mathematics at UCSD (2011).
 - M.S. in Computational Math from Stanford University (2013).
 - PhD in EECS from MIT (2020).
- Career Path
 - Worked at Xerox PARC from 2013-2015
 - Started Sony AI immediately after my PhD.
 - Current Position: Research Scientist
 - Details on Job
 - Adding explainability to AI agents.
 - Explanations for diagnosis and debugging.

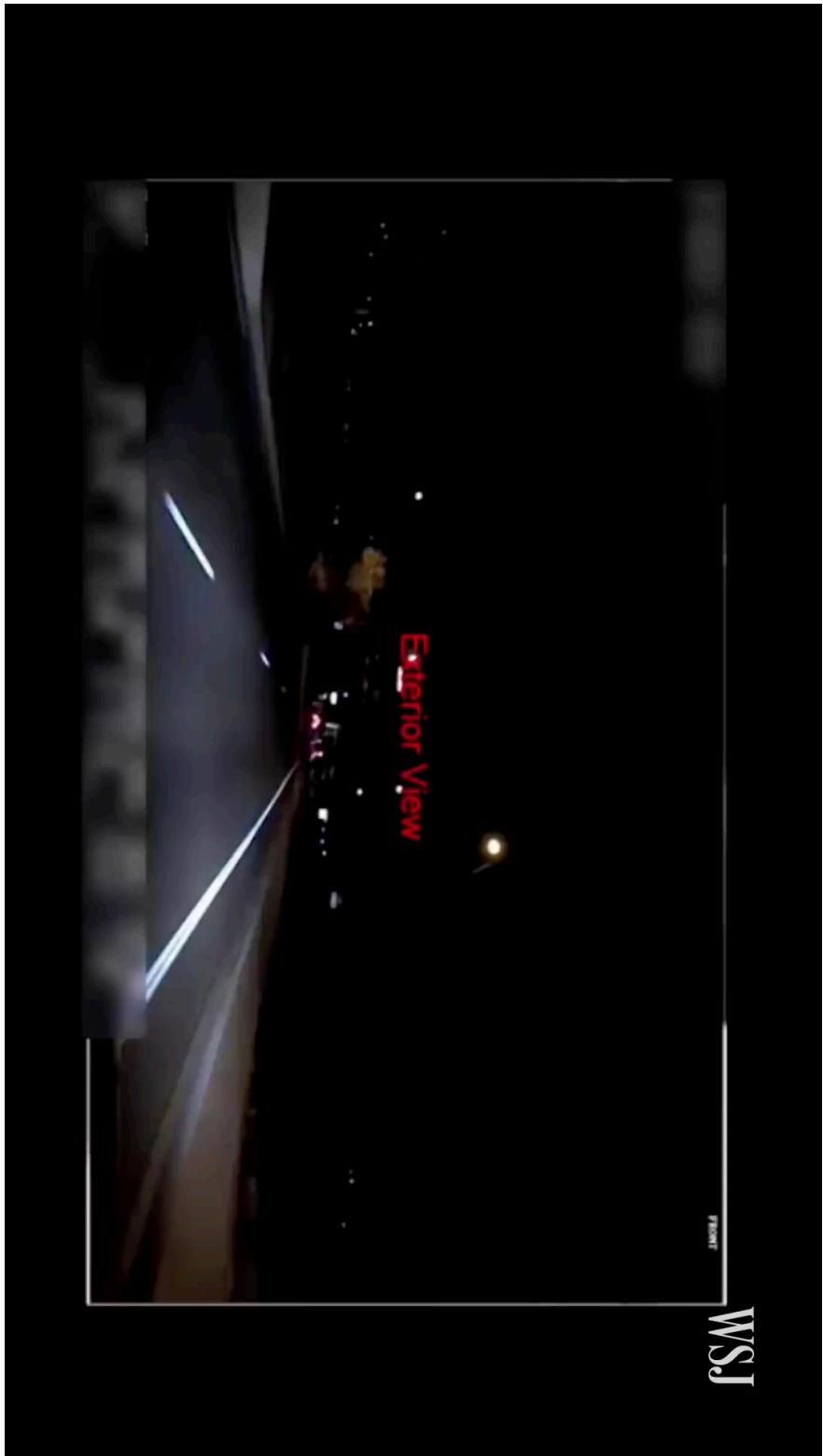
Local Sanity Checks



Synthesizer to reconcile
inconsistencies between parts.



A Deadly Crash

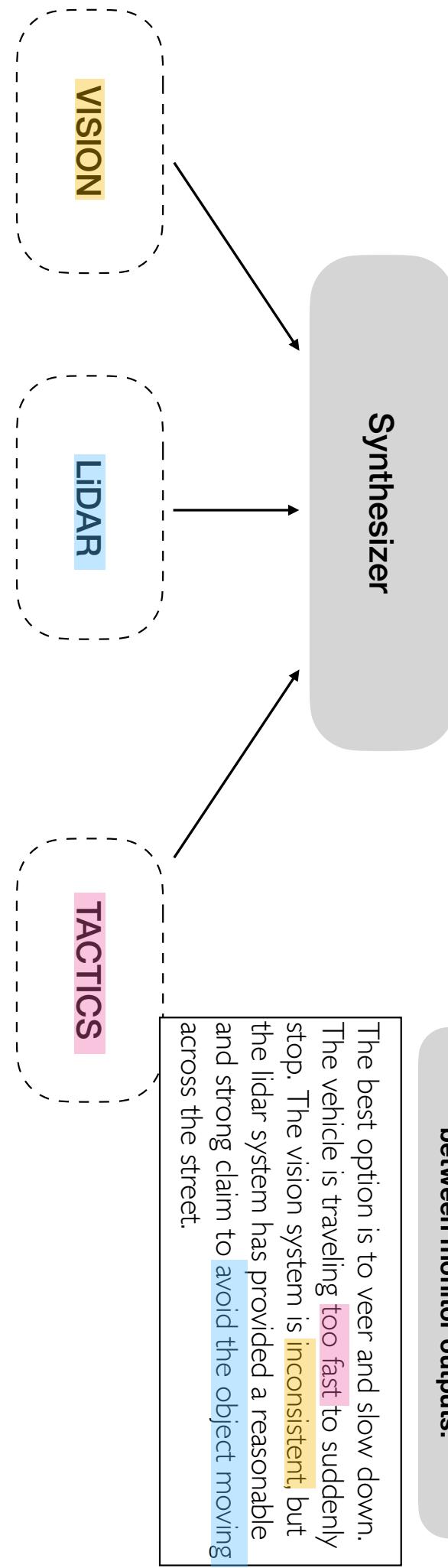


WSJ

Solution: Internal Communication

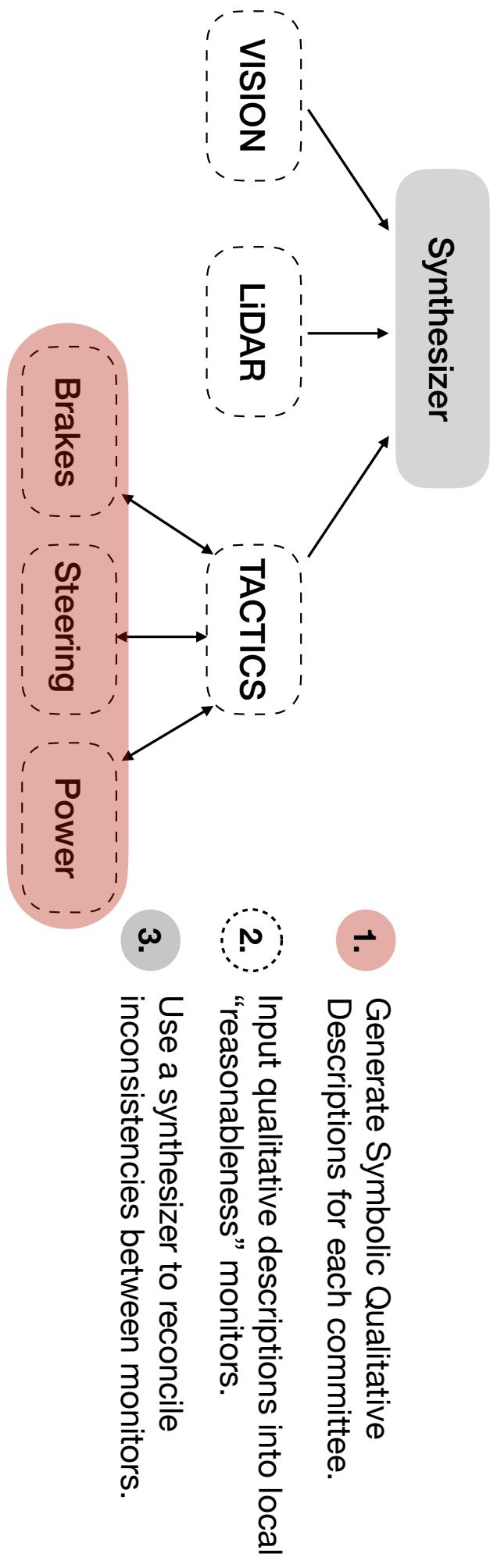
Anomaly Detection through Explanations

Synthesizer to reconcile inconsistencies
between monitor outputs.

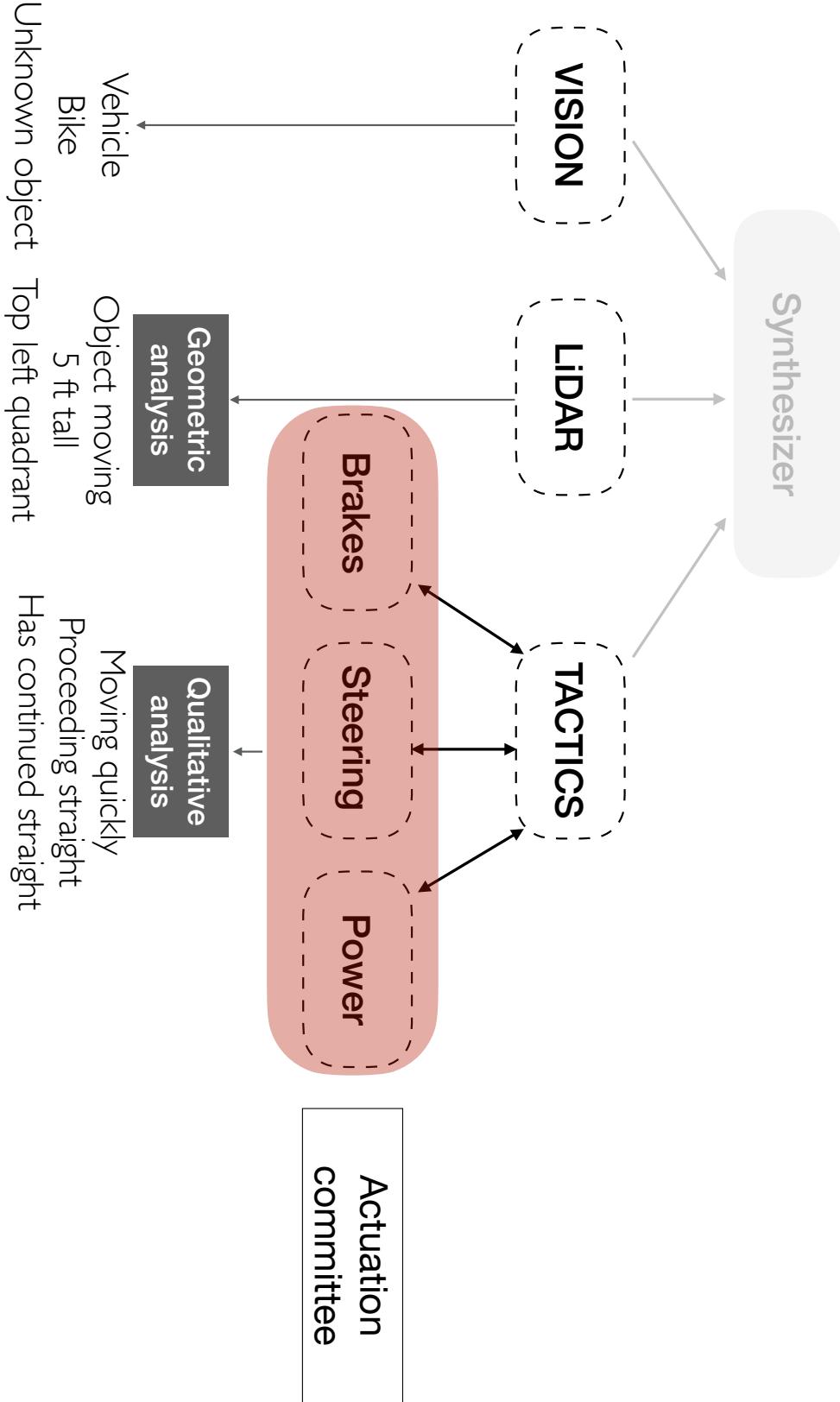


Anomaly Detection through Explanations

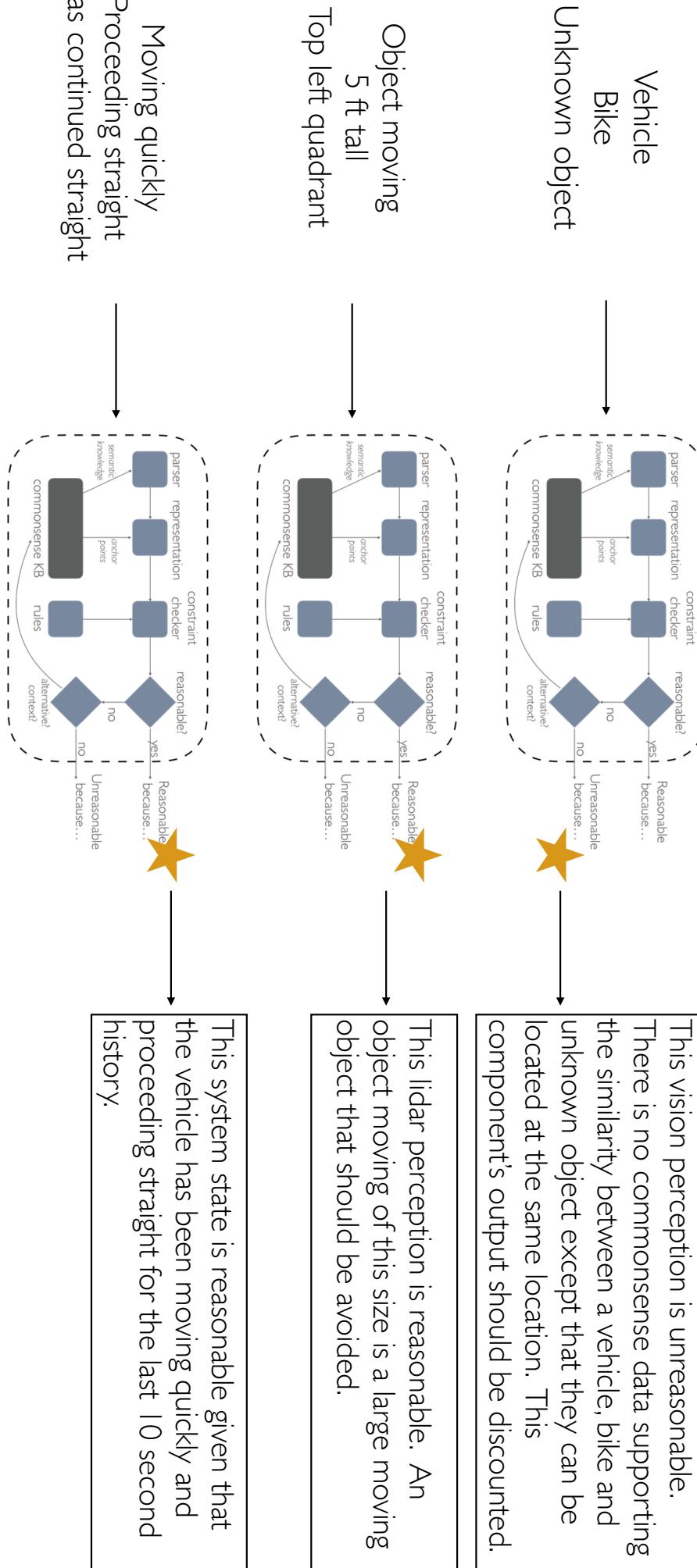
Reasoning in Three Steps



- 1.** Generate Symbolic Qualitative Descriptions for each committee.



2 Input qualitative descriptions into local “reasonableness” monitors.



3. Use a synthesizer to reconcile inconsistencies between monitors.

This vision perception is unreasonable. There is no commonsense data supporting the similarity between a vehicle, bike and unknown object except that they can be located at the same location. This component's output should be discounted.

This lidar perception is reasonable. An object moving of this size is a large moving object that should be avoided.

This system state is reasonable given that the vehicle has been moving quickly and proceeding straight for the last 10 second history.

Synthesizer

The best option is to veer and slow down. The vehicle is travelling too fast to suddenly stop. The vision system is inconsistent, but the lidar system has provided a reasonable and strong claim to avoid the object moving across the street.

Symbolic reasons

3.
Use a synthesizer to reconcile
inconsistencies between monitors.

```
(monitor, judgement, unreasonable)
(input, isType, labels)
(all_labels, inconsistent, negRel)
(isA, hasProperty, negRel)
...
(all_labels, notProperty, nearMiss)
(all_labels, locatedAt, consistent)
(monitor, recommend, discount)

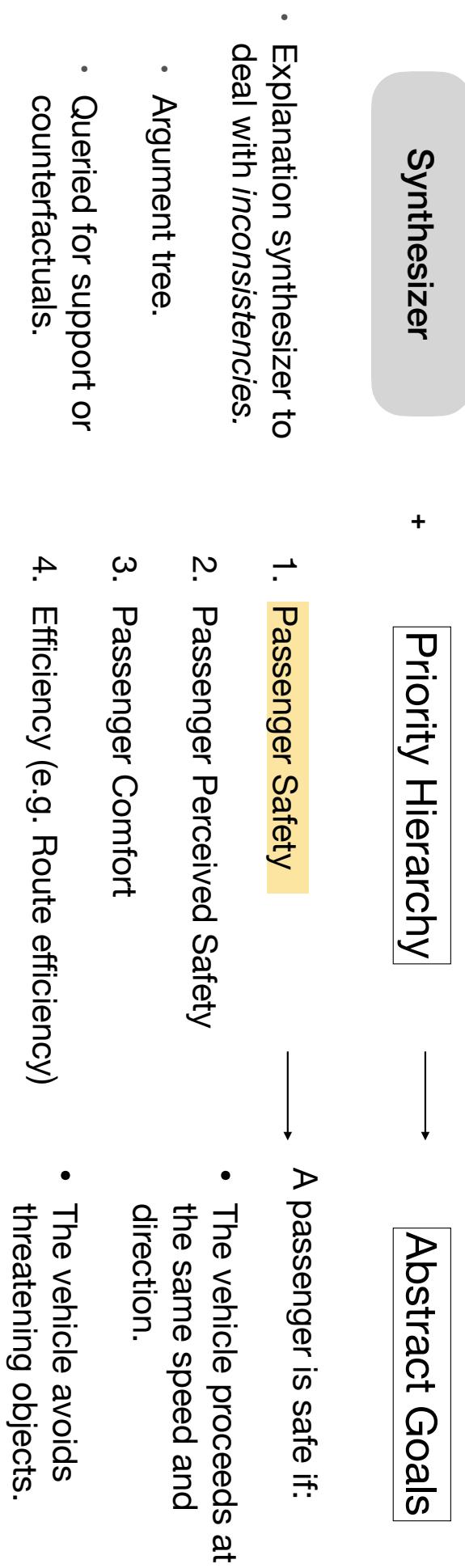
(monitor, judgement, reasonable)
(input_data, isType, sensor)
...
(input_data[4], hassize, large)
(input_data[4], IsA, large_object)
(input_data[4], moving, True)
(input_data[4], hasProperty, avoid)

(monitor, judgement, reasonable)
(input, isType, history)
(input_data, moving, True)
(input_data, direction, forward)
(input_data, speed, fast)
(input_data, consistent, true)
(monitor, recommend, proceed)
```

Synthesizer

The best option is to veer and slow down. The vehicle is travelling **too fast** to suddenly stop. The vision system is **inconsistent**, but the lidar system has provided a reasonable and strong claim to **avoid the object moving** across the street.

3. Use a synthesizer to reconcile inconsistencies between monitors.



3.

Use a synthesizer to reconcile inconsistencies between monitors.

```
(monitor, judgement, unreasonable)
(input, isType, labels)
(all_labels, inconsistent, negRel)
(isA, hasProperty, negRel)
...
(all_labels, notProperty, nearMiss)
(all_labels, locatedAt, consistent)
(monitor, recommend, discount)
...
(monitor, judgement, reasonable)
(input, isType, sensor)
...
(input_data[4], bassize, large)
(input_data[4], ISA, large_object)
(input_data[4], moving, True)
(input_data[4], hasProperty, avoid)
...
(monitor, recommend, avoid)
...
(monitor, judgement, reasonable)
(input, isType, history)
(input_data, moving, True)
(input_data, direction, forward)
(input_data, speed, fast)
(input_data, consistent, True)
(monitor, recommend, proceed)
```

Abstract Goal Tree

```
'passenger is safe',
AND(
  'safe transitions',
  NOT('threatening objects'))!
```

The best option is to veer and slow down. The vehicle is traveling **too fast** to suddenly stop. The vision system is **inconsistent**, but the lidar system has provided a reasonable and strong claim to **avoid the object moving** across the street.

Evaluation in Simulation

