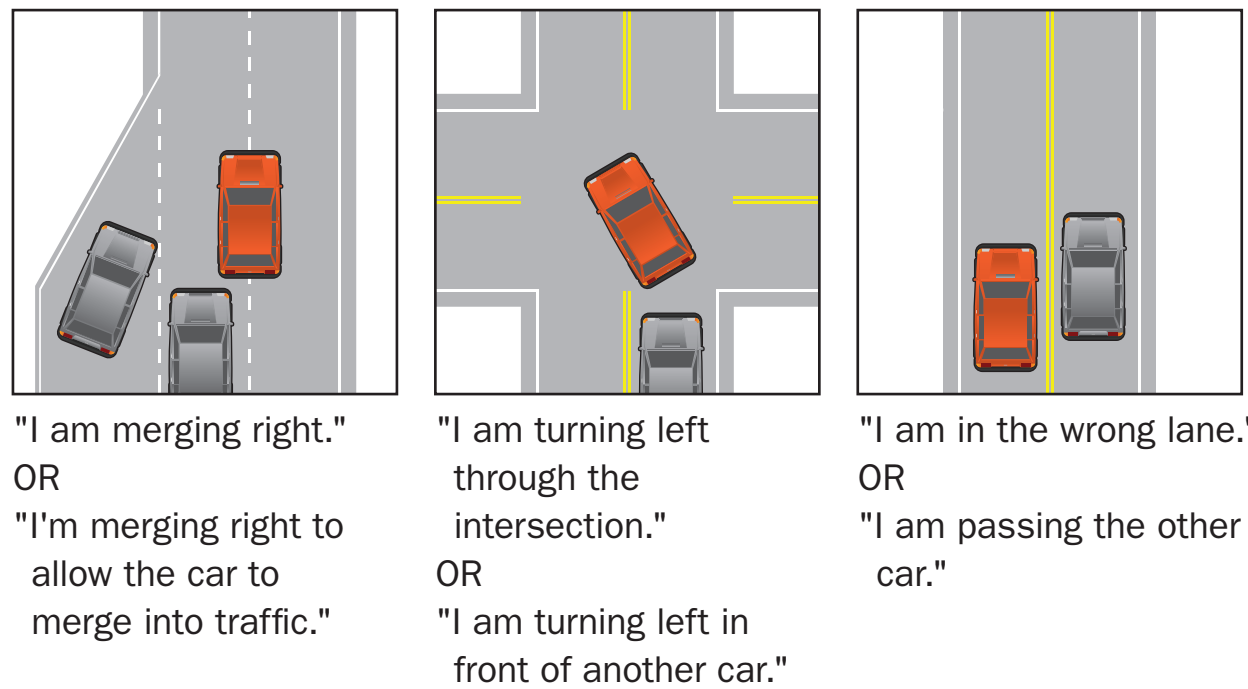# Why did the robot do that?

Robots are increasingly being utilized in important tasks such as search and rescue operations. However, their behaviors are often hard to distinguish and understand, leading to users' mistrust and often abandonment of very useful tools. We are developing algorithms for robots to automatically explain their behaviors to users and are demonstrating that these explanations improve users' trust and acceptance of them compared to robots that do not explain themselves.

? How can we capture diverse sets of user preferences for what the robot explains?

Representing preferences: We have developed a set of parameters that allow us to capture preferences such as level of abstraction and length and automatically generate different explanations based on those preferences.

User Interaction: The user can query the robot for more or different information if their preferences change or they want to dig deeper into the explanation.

? How can we generate explanations of a diverse set of robots, sensors, actions, and tasks?

Our Methodology

We first poll many people to capture many different ways to explain example robot behaviors.

"I am merging right."
OR
"I'm merging right to allow the car to merge into traffic."

"I am turning left through the intersection."
OR
"I am turning left in front of another car."

"I am in the wrong lane."
OR
"I am passing the other car."

Then, we poll a new set of people to measure which words and explanations are best.

By analyzing the ranked explanations we can capture patterns of language that the robots should use in their explanations.
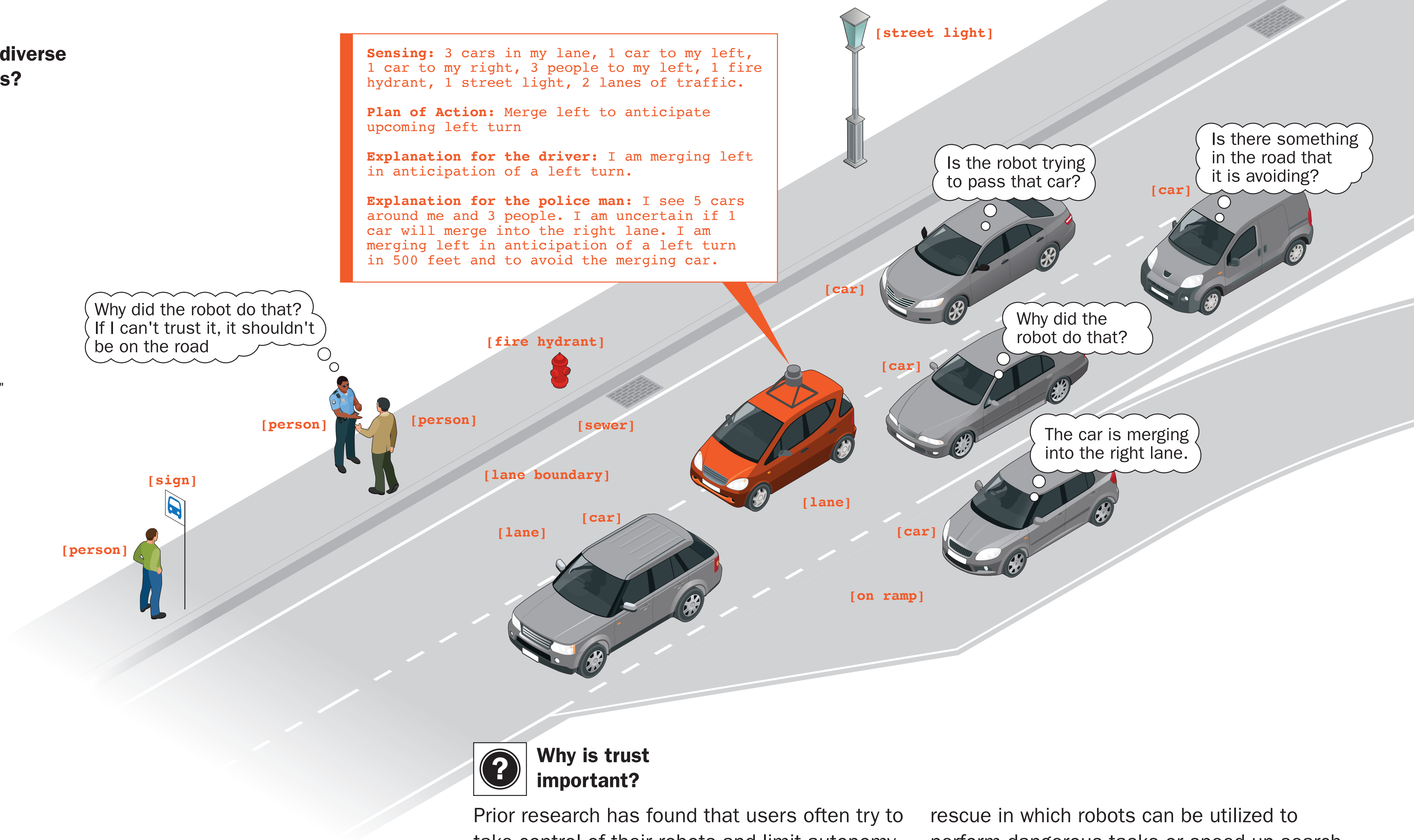
Order of importance:

1. Describing action
2. Describing immediate scene
3. Describing surrounding scene
4. Describing uncertainty in scene

Sensing: 3 cars in my lane, 1 car to my left, 1 car to my right, 3 people to my left, 1 fire hydrant, 1 street light, 2 lanes of traffic.

Plan of Action: Merge left to anticipate upcoming left turn

Explanation for the driver: I am merging left in anticipation of a left turn.

Explanation for the police man: I see 5 cars around me and 3 people. I am uncertain if 1 car will merge into the right lane. I am merging left in anticipation of a left turn in 500 feet and to avoid the merging car.

[street light]

Is the robot trying to pass that car?

Is there something in the road that it is avoiding?

[car]

[car]

Why did the robot do that?

[car]

The car is merging into the right lane.

Why did the robot do that? If I can't trust it, it shouldn't be on the road

[fire hydrant]

[person]   [person]

[sewer]

[lane boundary]

[sign]

[lane]

[car]

[lane]

[car]

[on ramp]

[person]

? Why is trust important?

Prior research has found that users often try to take control of their robots and limit autonomy when they lose trust in them, taking the user focus off the task at hand. Especially in time-sensitive applications like search and rescue in which robots can be utilized to perform dangerous tasks or speed up search tasks, we cannot afford to have first-responders lose trust or even stop using robots.

Contact: Stephanie Rosenthal, PhD | srosenthal@sei.cmu.edu