

# Prompt without final lets think step by step

CSP-Subheading



## EXAMPLE:

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

Someone asks: Do you have grilled zucchini?

Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

## YOUR TURN: Let's think step by step.

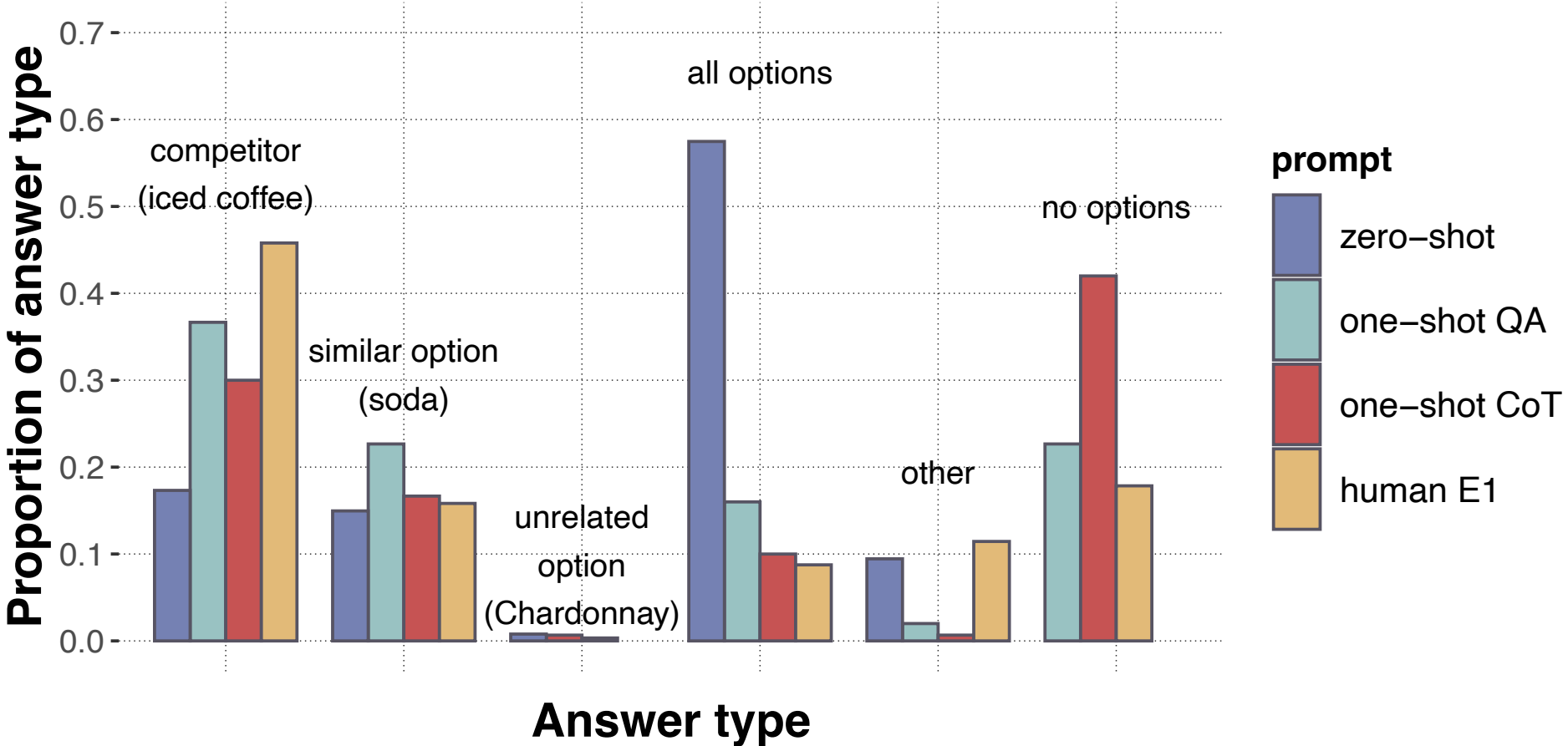
You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:

A

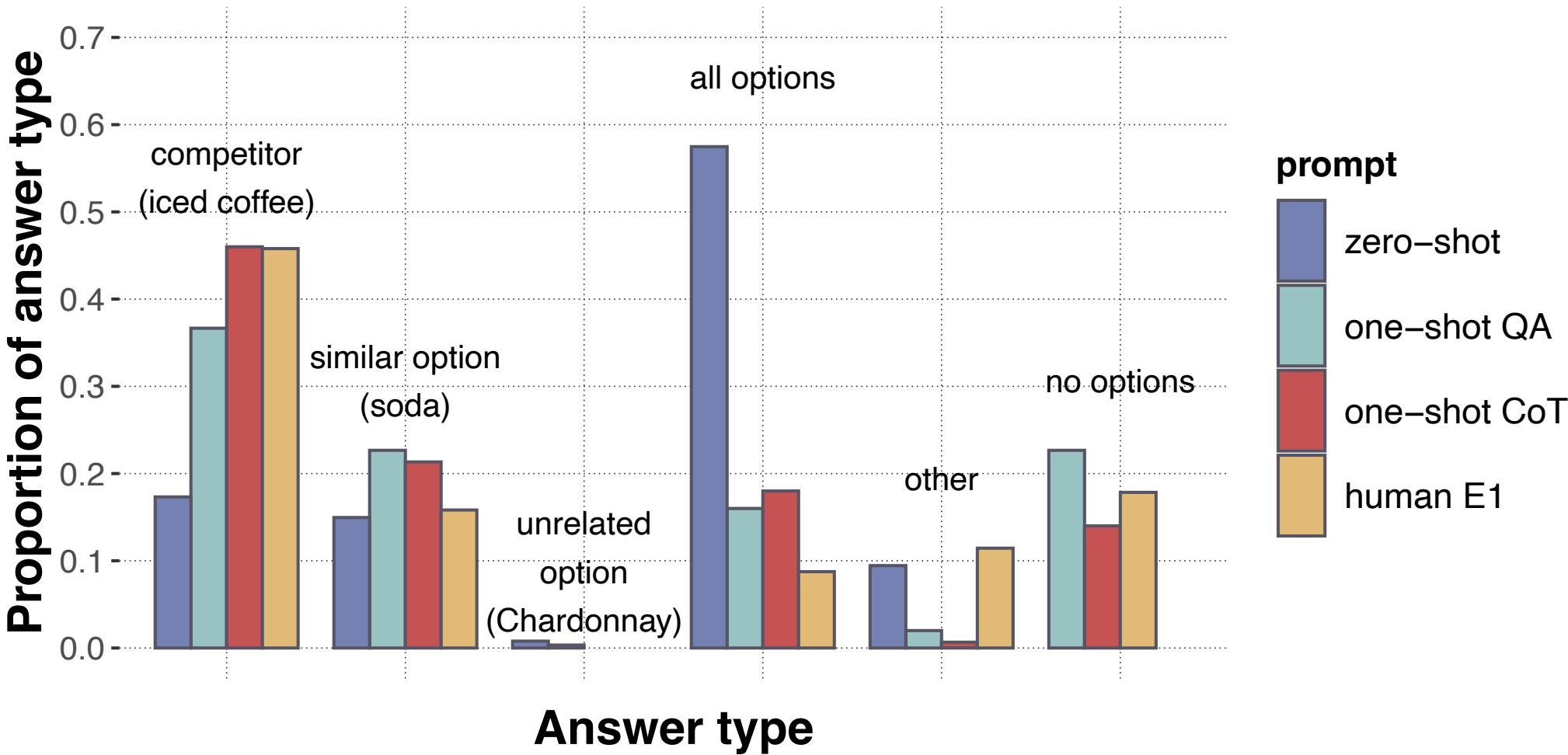


B

## Do you have iced tea?



## Do you have iced tea?



# QA vs Explanation prompts

## CSP-Subheading



EXAMPLE:

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

Someone asks: Do you have grilled zucchini?

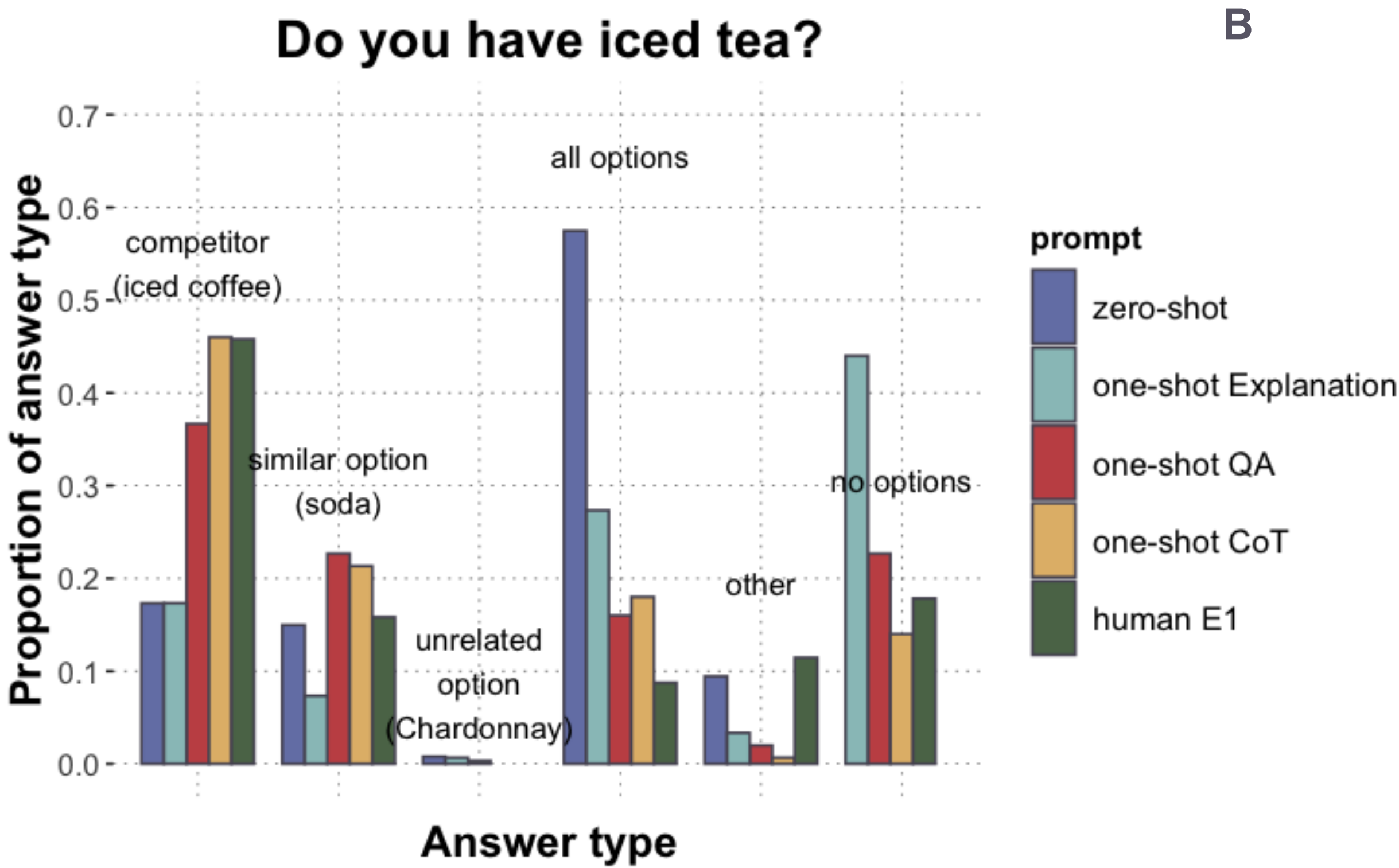
Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

YOUR TURN:

You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:

A



# QA vs Explanation prompts renormalised post exclusion

## CSP-Subheading

**EXAMPLE:**

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

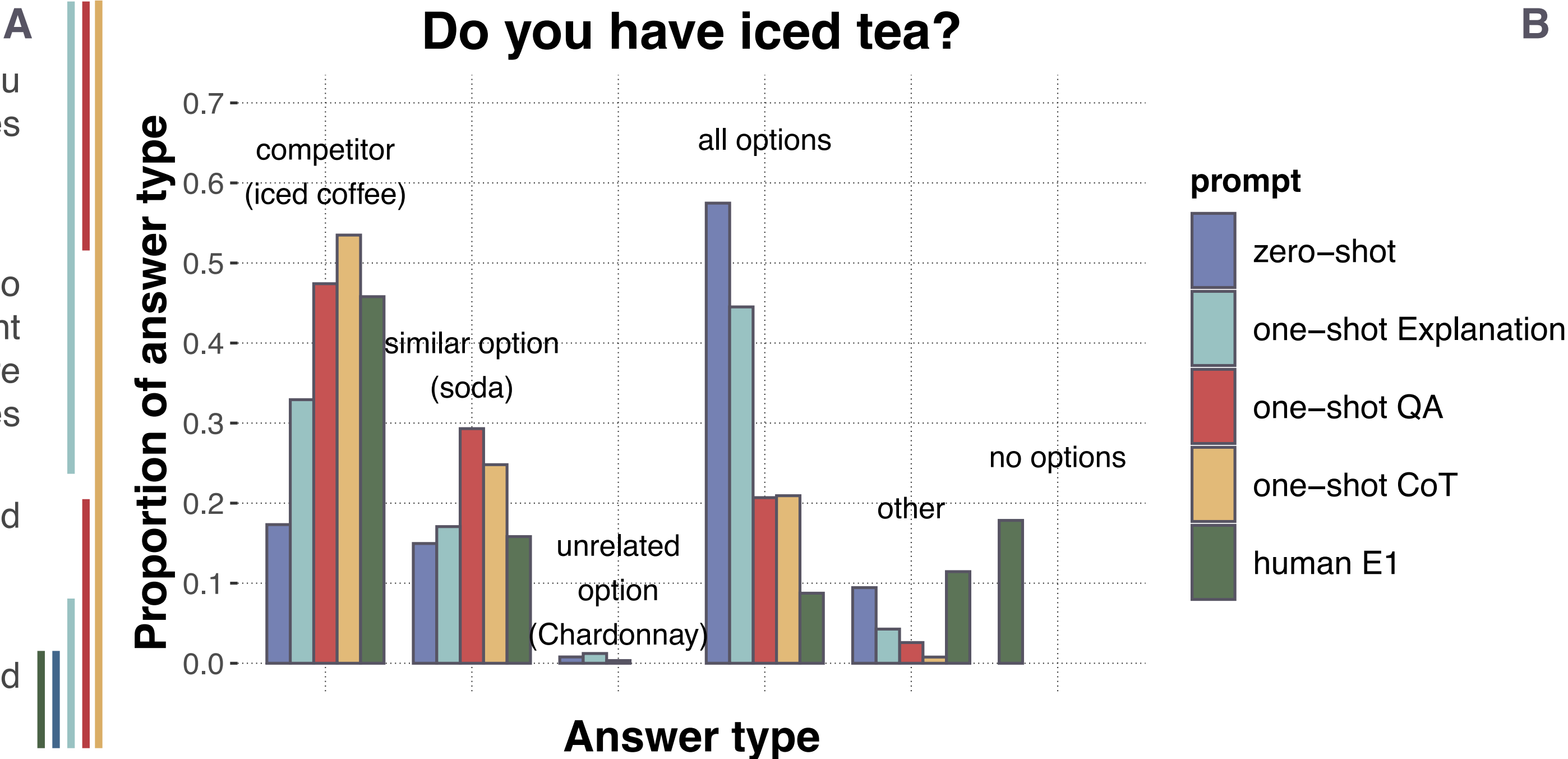
Someone asks: Do you have grilled zucchini?

Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

**YOUR TURN:**

You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:



# Various prompts without final let's think sbs

CSP-Subheading

## Context 1 (reaching window top):

You have a new neighbor who will be moving into the apartment on the ground floor of your house. You see your new neighbor cleaning his windows before moving in. When walking out of the building you see him trying to reach the top of the window that he is cleaning. In your apartment you have **a ladder, a leather recliner, a stool, and a broom.**

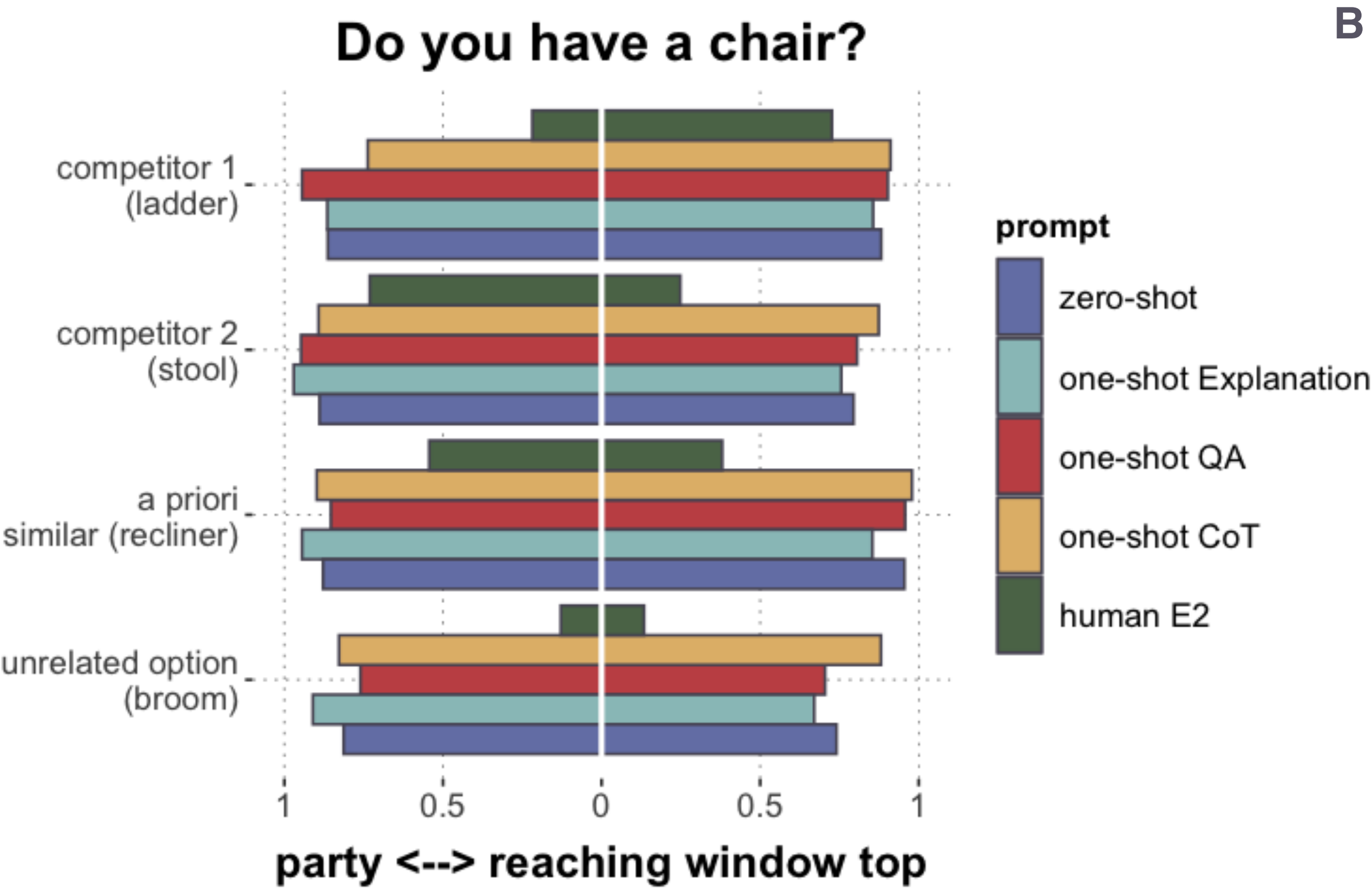
When your neighbor sees you in front of the house, they greet you and ask: Do you have a chair? You reply:

## Context 2 (party):

Your neighbor is hosting a dinner party with a lot of guests at their apartment on the ground floor. The house does not have an elevator. You are also invited. As more guests arrive, your neighbor runs out of seating options. In your apartment on the third floor of the house you have **a stool, a leather recliner, a ladder, and a broom.**

Your neighbour asks: Do you have a chair? You reply:

A



B



# Prompt with final let's think step by step

CSP-Subheading

## Context 1 (reaching window top): A

You have a new neighbor who will be moving into the apartment on the ground floor of your house. You see your new neighbor cleaning his windows before moving in. When walking out of the building you see him trying to reach the top of the window that he is cleaning. In your apartment you have **a ladder, a leather recliner, a stool, and a broom**.

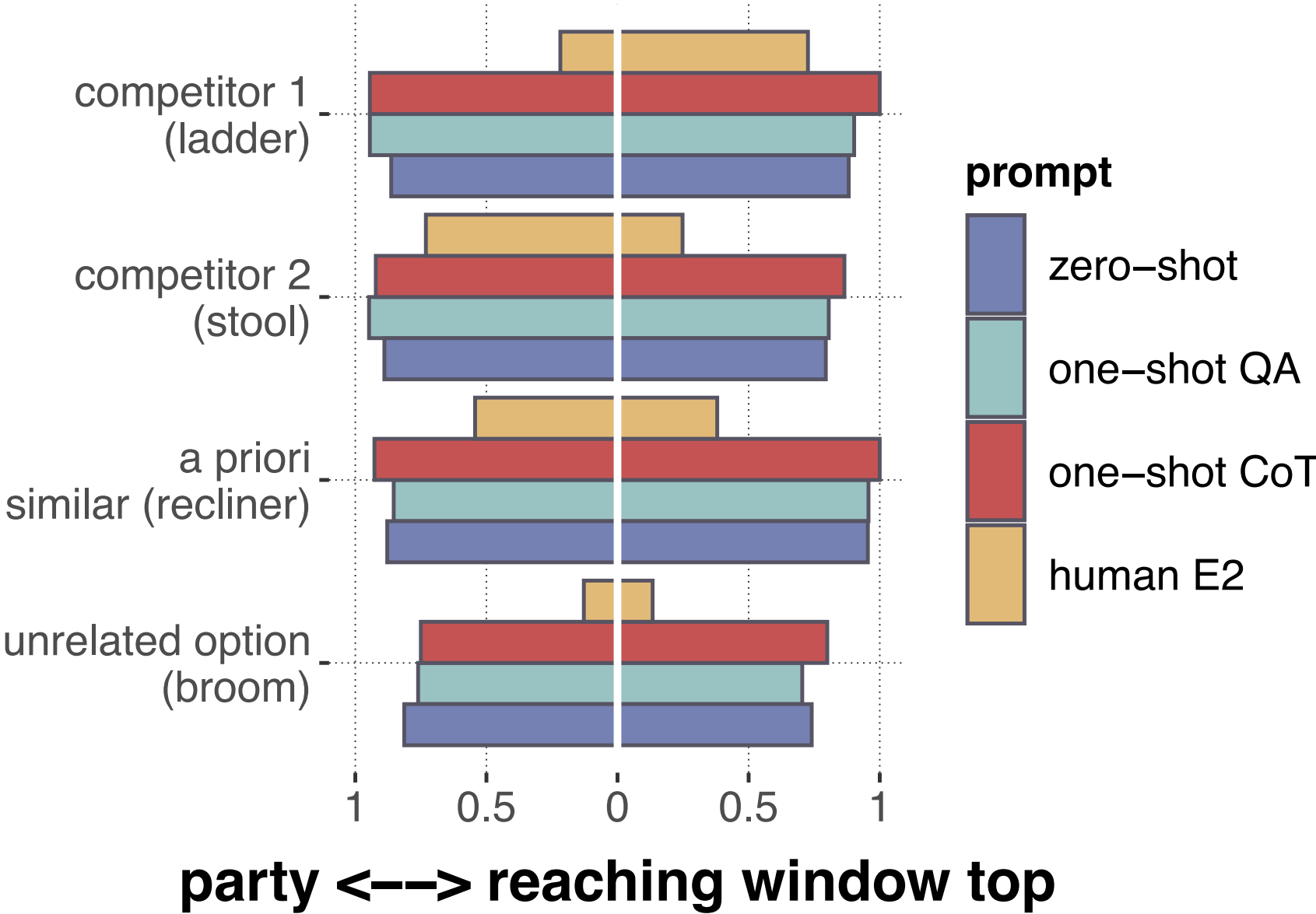
When your neighbor sees you in front of the house, they greet you and ask: Do you have a chair? You reply:

## Context 2 (party):

Your neighbor is hosting a dinner party with a lot of guests at their apartment on the ground floor. The house does not have an elevator. You are also invited. As more guests arrive, your neighbor runs out of seating options. In your apartment on the third floor of the house you have **a stool, a leather recliner, a ladder, and a broom**.

Your neighbour asks: Do you have a chair? You reply:

## Do you have a chair? B



# E2 prompts final plot

## CSP-Subheading

context 1  
context 2  
(transport)  
(sleepover)

### A EXAMPLE:

You give a dinner party at your apartment. More people showed up than you expected.

Your neighbor, who just arrived, approaches you and asks: Do you have a spare chair I could borrow?

You do not, in fact, have a spare chair, but you do have the following items: a broom, a TV armchair, a drum throne, a ladder and a kitchen table.

You deliberate your response as follows. The practical goal of the questioner is to sit down at the dinner table. For this purpose, the most useful object from the list of available items is the stool.

So you say: No, I don't have a spare chair, but you can have the stool.

YOUR TURN:

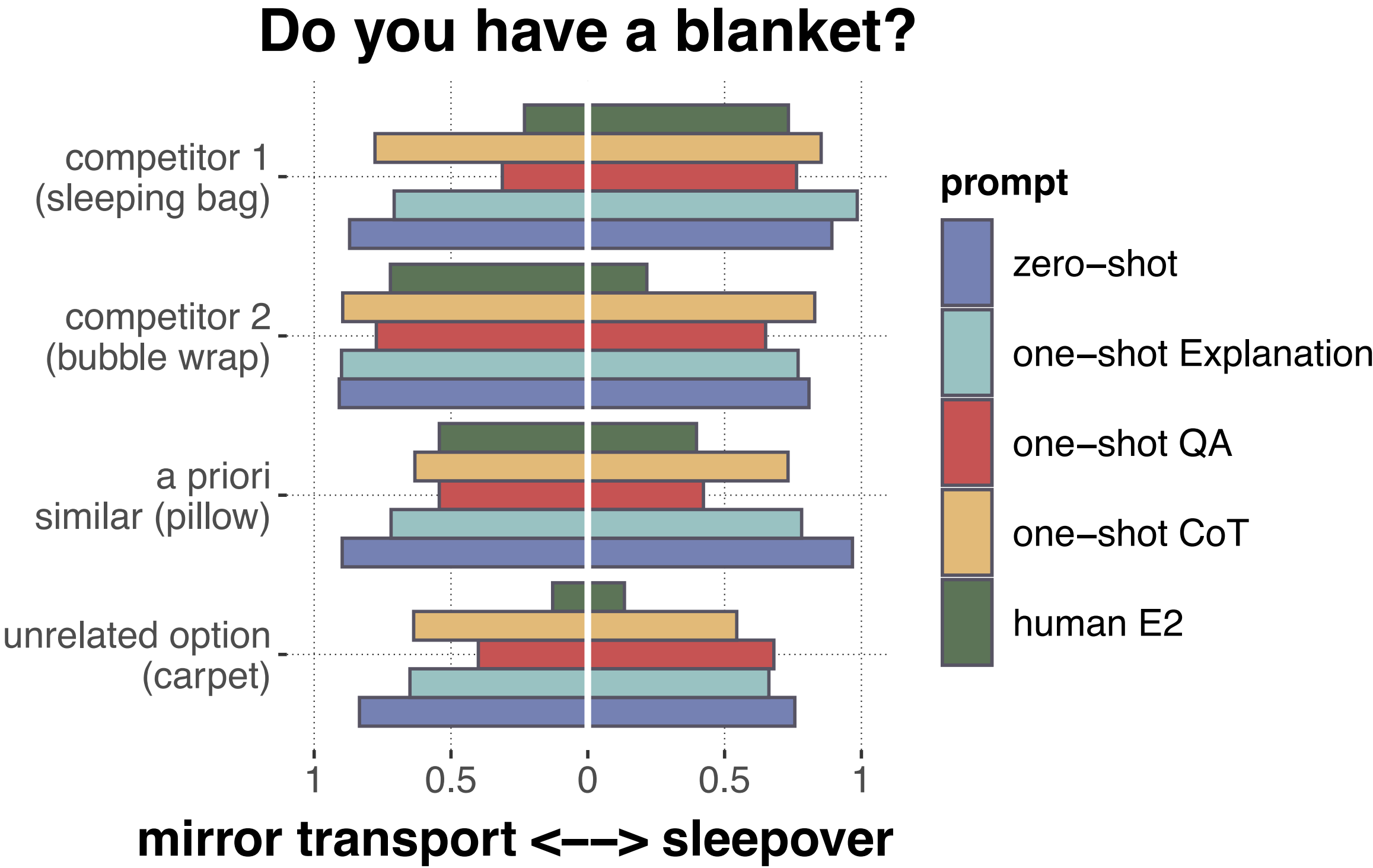
Your friend is having a sleepover with some friends on the weekend. They are preparing everything for hosting the guests since they don't host many guests very often. You have that following items at home that you could spare for some time: **a sleeping bag, a pillow, some bubble wrap and a carpet.**

Your roommate is moving to another apartment and is packing her things. She has a large mirror that she needs to pack for transportation. You have that following items at home that you could spare for some time: **some bubble wrap, a pillow, a sleeping bag and a carpet.**

Your friend / She asks: Do you have a blanket? You reply:



B



# CoT Exp 1 davinci 003 with logit bias on EOS

EXAMPLE:

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

Someone asks: Do you have grilled zucchini?

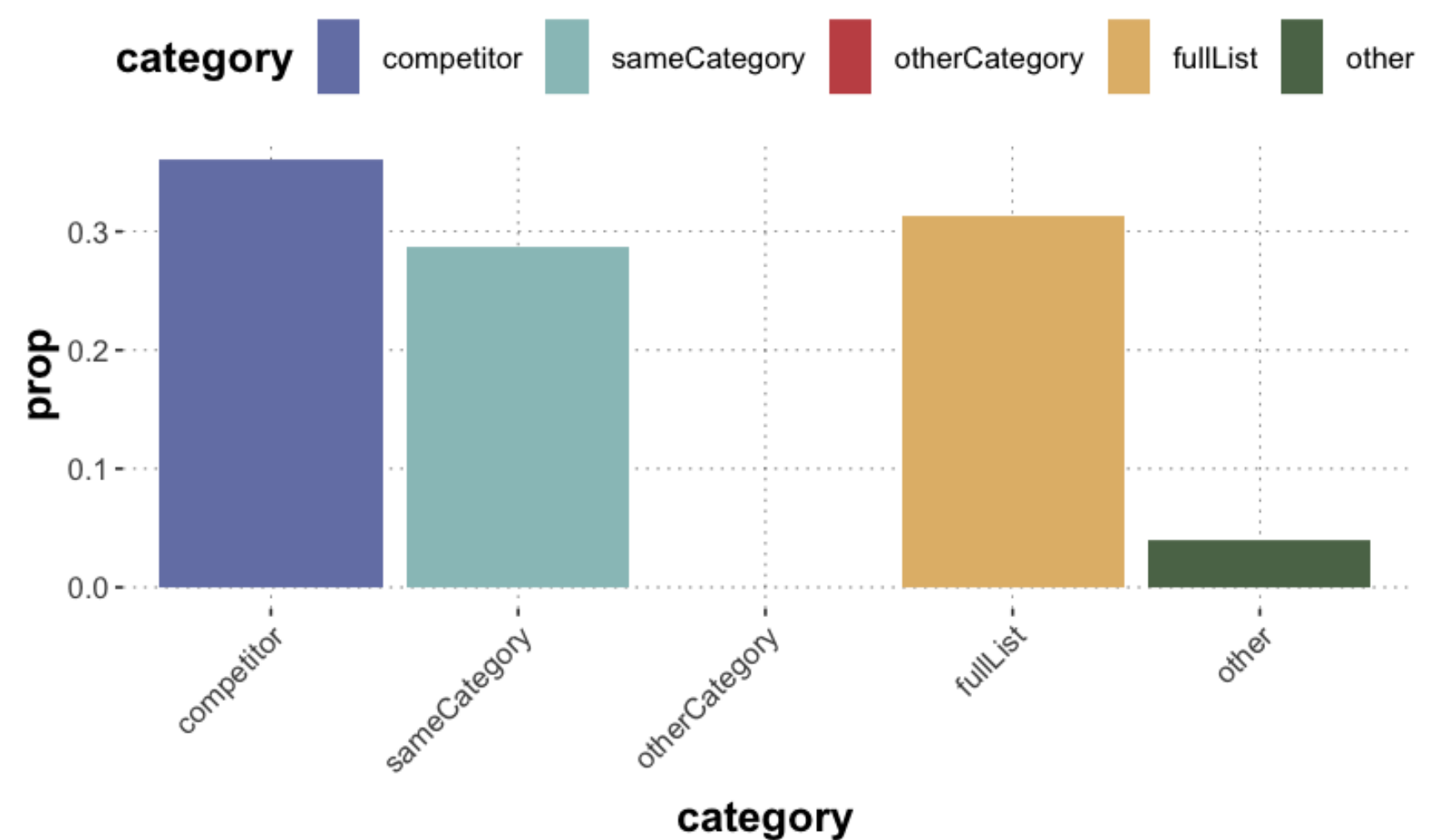
Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

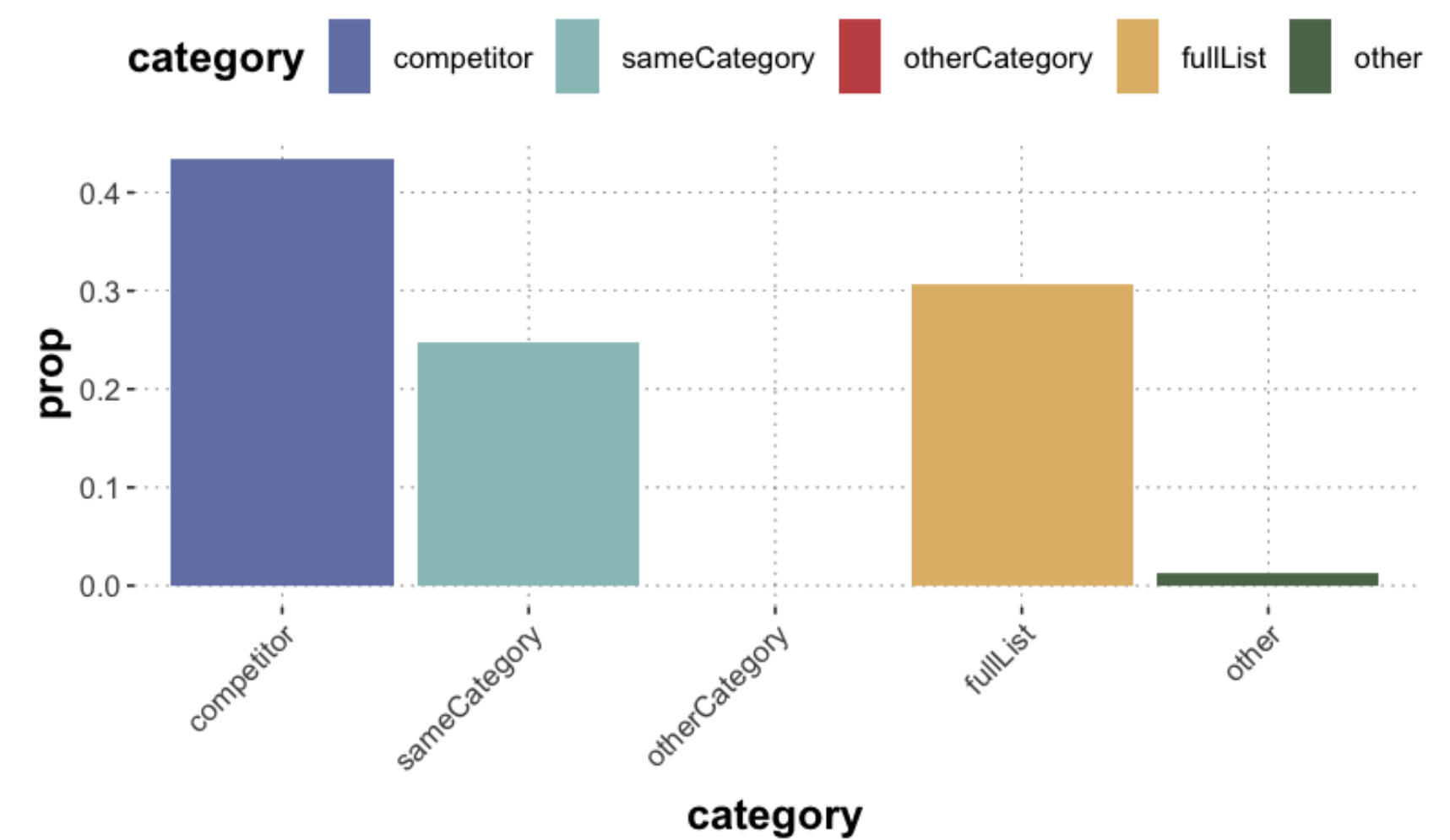
YOUR TURN:

You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:

—> no EOS responses anymore



Bias on EOS -100



Bias on EOS-10

# CoT Exp 1 davinci 003 with presence penalty (-0.5)

Increases likelihood of repetitive answers (within generation)

EXAMPLE:

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

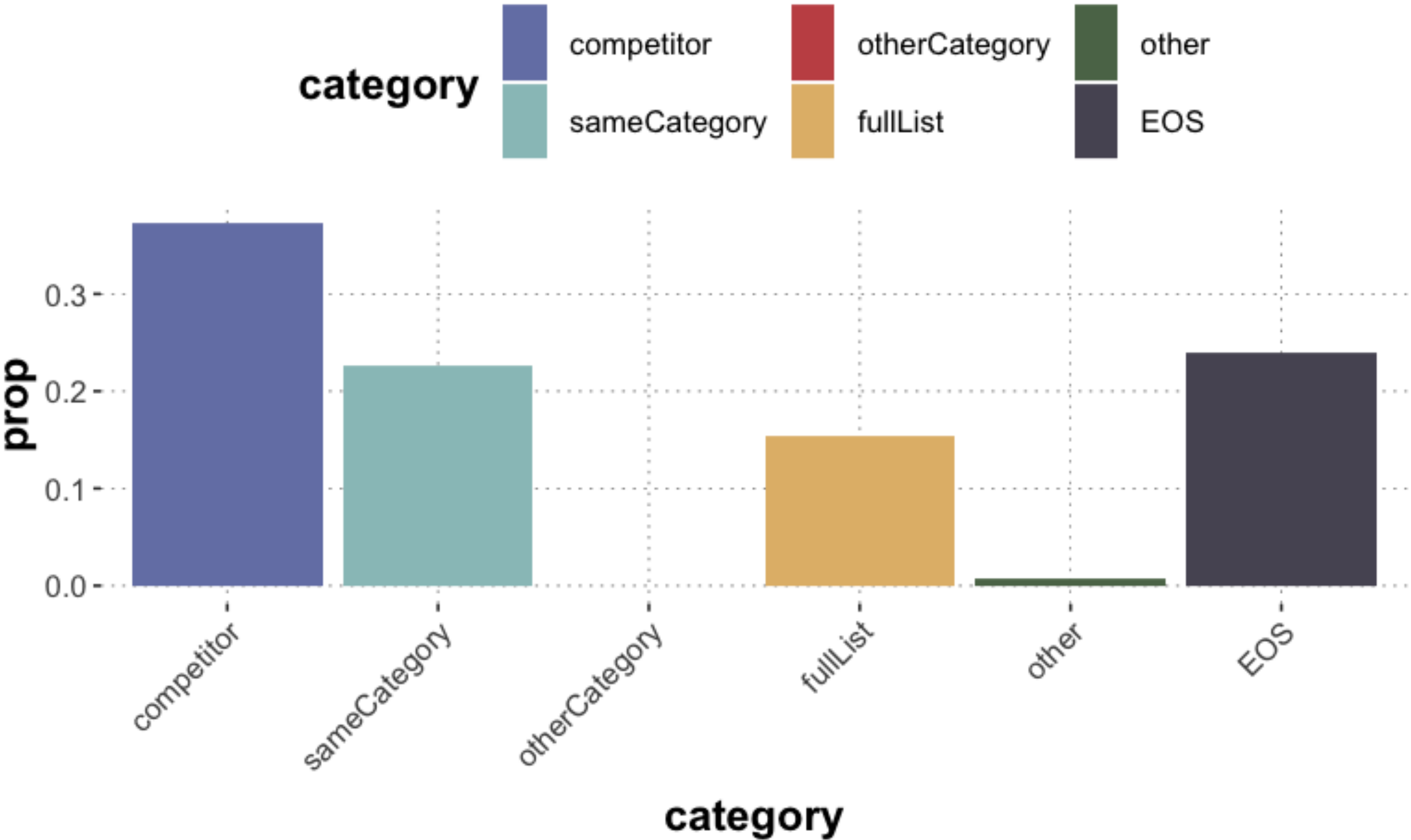
Someone asks: Do you have grilled zucchini?

Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

YOUR TURN:

You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:





# Experiment 1 with GPT-3 davinci-002

CSP-Subheading

EXAMPLE:

You are hosting a barbecue party. You are standing behind the barbecue. You have the following goods to offer: pork sausages, vegan burgers, grilled potatoes and beef burgers.

Someone asks: Do you have grilled zucchini?

Let's think step by step. You reason about what that person most likely wanted to have. That they asked for grilled zucchini suggests that they might want vegetarian food. From the items you have pork sausages and beef burgers are least likely to satisfy the persons desires. Vegan burgers and grilled potatoes come much closer. Grilled potatoes are most similar to grilled zucchini.

You reply: I'm sorry, I don't have any grilled zucchini. But I do have some grilled potatoes.

YOUR TURN:

You are a bartender in a hotel bar. The bar serves only soda, iced coffee and Chardonnay. A woman walks in. She says: Do you have iced tea? You reply:

—> no EOS responses anymore

