

GPT-3
(Log-odds)

-10

-5

0

5

10

0.09

0.06

0.03

0.00

Human
(Proportion)

0.00

0.25

0.50

0.75

1.00

6

4

2

0

Bias toward Start location

Knowledge State



False Belief



True Belief

