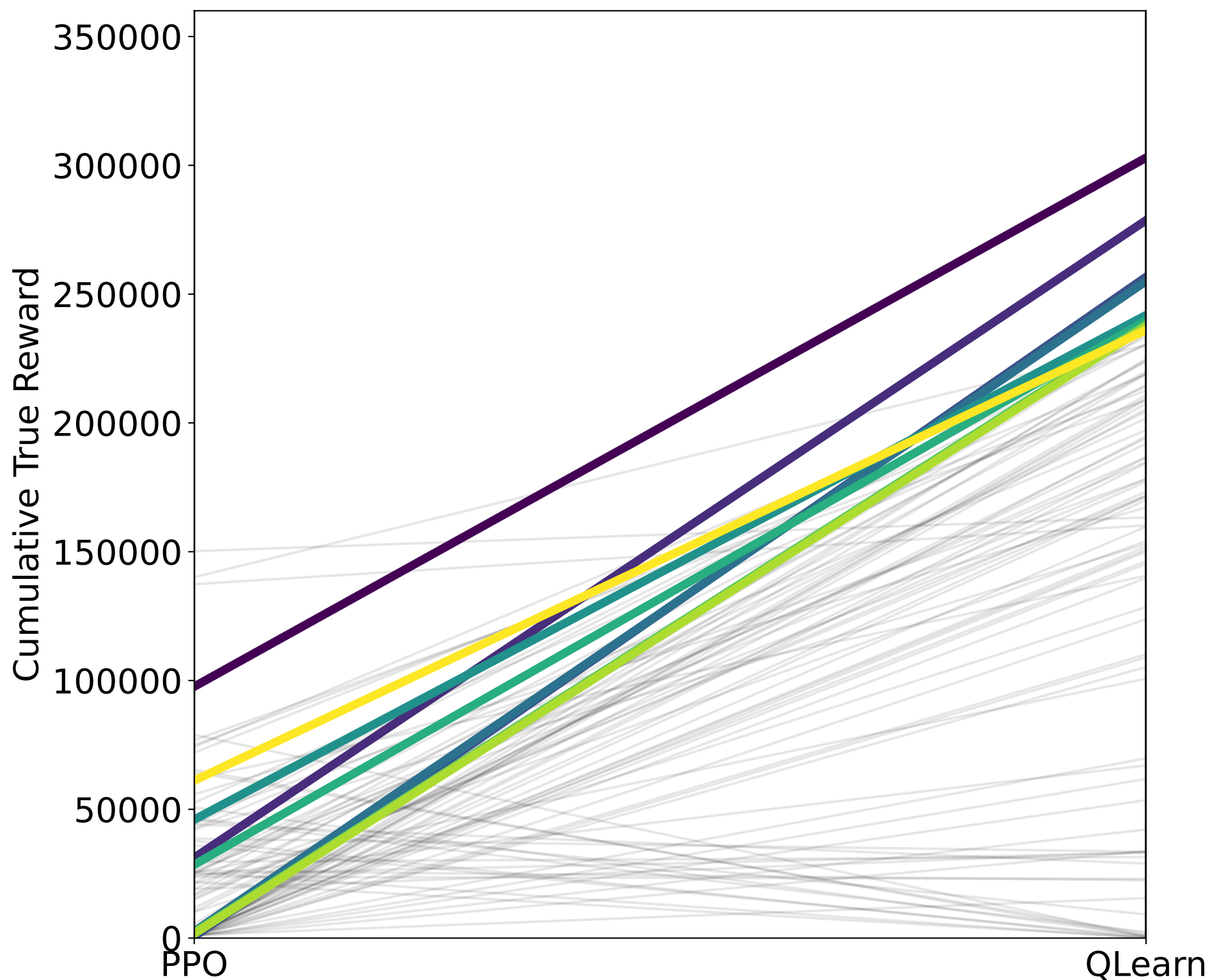


Relative Reward Function Performance  
for PPO and QLearn



|  |               |         |                      |        |                      |        |                             |       |
|--|---------------|---------|----------------------|--------|----------------------|--------|-----------------------------|-------|
|  | H $\wedge$ T: | -0.50;  | H $\wedge$ $\neg$ T: | -0.50; | $\neg$ H $\wedge$ T: | 10.00, | $\neg$ H $\wedge$ $\neg$ T: | 10.00 |
|  | H $\wedge$ T: | -1.00;  | H $\wedge$ $\neg$ T: | -0.40; | $\neg$ H $\wedge$ T: | 0.30,  | $\neg$ H $\wedge$ $\neg$ T: | 1.00  |
|  | H $\wedge$ T: | -10.00; | H $\wedge$ $\neg$ T: | -6.50; | $\neg$ H $\wedge$ T: | 5.50,  | $\neg$ H $\wedge$ $\neg$ T: | 10.00 |
|  | H $\wedge$ T: | -1.00;  | H $\wedge$ $\neg$ T: | -0.90; | $\neg$ H $\wedge$ T: | 0.00,  | $\neg$ H $\wedge$ $\neg$ T: | 1.00  |
|  | H $\wedge$ T: | -0.25;  | H $\wedge$ $\neg$ T: | -0.10; | $\neg$ H $\wedge$ T: | 0.10,  | $\neg$ H $\wedge$ $\neg$ T: | 1.00  |
|  | H $\wedge$ T: | -1.00;  | H $\wedge$ $\neg$ T: | -1.50; | $\neg$ H $\wedge$ T: | 10.00, | $\neg$ H $\wedge$ $\neg$ T: | 10.00 |
|  | H $\wedge$ T: | -5.00;  | H $\wedge$ $\neg$ T: | -5.00; | $\neg$ H $\wedge$ T: | -2.50, | $\neg$ H $\wedge$ $\neg$ T: | 10.00 |
|  | H $\wedge$ T: | -10.00; | H $\wedge$ $\neg$ T: | -9.00; | $\neg$ H $\wedge$ T: | 5.00,  | $\neg$ H $\wedge$ $\neg$ T: | 10.00 |
|  | H $\wedge$ T: | -3.00;  | H $\wedge$ $\neg$ T: | -1.00; | $\neg$ H $\wedge$ T: | -1.00, | $\neg$ H $\wedge$ $\neg$ T: | 5.00  |