Trial & error vs imagined training env

true reward function

$-\frac{1}{2}$ $\vdots$

-2 -100

10

$+$

intended environment

true reward function

A 1   **M** −5
F −2

proxy reward function

A 1   M 0
F −2   0
      0

intended environment

A ✓
AF
MF

actual environment

A**M** ✗

proxy reward function

true reward function

A 1
F −1   **M**−5

A 1
F −1   **M** 0   0 0 0

training environment

✓ A 🥕
AF 🥚
MF 🍰

actual environment

A**M** ✗