

Entity Prediction

Overall, when model predicts a event change, 85% of the time it also predicts some entity change

- When a step
 - Has event prediction and actual event: 86% of the prediction has predicted entity state
 - Has event prediction but no actual event: 79% of the prediction has predicted entity state
 - Does not have event prediction but has actual event: 69% of the prediction has predicted entity state
- **73% of event predictions are paired with an (entity, state) prediction.**
 - Some steps contain sufficient information to give a concrete answer to the event. This is probably the reason why Codex doesn't produce entity state.
 - Example:

```
Goal: (Install a game on PS5)
Current Step: (Download and launch the game to start playing. )
True Events:
('I run the game.', 'more likely')
Predicted Events:
('I run the game.', 'more likely')
True Entities:
('the game', 'installed', 'more likely')

Goal: (Fold a shirt)
Current Step: (Repeat folding the shirt in half to reduce its size. )
True Events:
('The shirt takes up less space in my wardrobe.', 'more likely')
Predicted Events:
('The shirt takes up less space in my wardrobe.', 'more likely')
True Entities:
('shirt', 'reduced in size', 'more likely')
```

Codex doesn't tend to produce more than 1 state changes:

- Average number of predicted entity states on key steps: 1.08 (6)
- Average number of ground truth entity states on key steps: 1.21 (15)