

Neutral  
Stability

Contradiction  
Stability

Entailment  
Stability

(a)

(b)

(c)

Frequency

Frequency

Frequency

0.0 0.2 0.4 0.6 0.8 1.0

0.2 0.4 0.6 0.8 1.0

0.0 0.2 0.4 0.6 0.8 1.0

- Original
- Perturbed
- Optimized
- Ground Truth

- Original
- Perturbed
- Optimized
- Ground Truth

Neutral

Neutral

The original prediction is a boundary case

( $\mathbf{e}_1$ ) Prediction is altered by replacing **pile** with **heap**

Contradiction

(d)

Entailment

Contradiction

(e)

Entailment