# Interests and Evidence

Brian Weatherson, February 26, 2017

## Games

| | Predict 1 Box | Predict 2 Boxes |
|---|---|---|
| Choose 1 Box | 1000, 1 | 0,0 |
| Choose 2 Boxes | 1001, 0 | 1, 1 |

*Game 1: Newcomb*

| | $p \in E$ | $p \notin E$ |
|---|---|---|
| Take the bet | **1, 1** | -9.1, 0 |
| Decline the bet | 0, 0 | **0, 1** |

*Game 2: Interpretation*

| | $a$ | $b$ |
|---|---|---|
| $A$ | 5, 5 | 0, 4 |
| $B$ | 4, 0 | 2, 2 |

*Game 3: Stag Hunt*

| | $a$ | $b$ |
|---|---|---|
| $A$ | 4, 4 | 0, 4 |
| $B$ | 4, 0 | 2, 2 |

*Game 4: Pareto Dominant is Weakly Dominated*

| | $p \in E$ | $p \notin E$ |
|---|---|---|
| Take the bet | 1, 1 | -0.6, 0 |
| Decline the bet | 0, 0 | 0, 1 |

*Game 5: Interpretation Game with Low Odds*

## Rules for Interpretation Game

There are two players:

1. Human (on row)
2. Radical Interpreter (on column)

Here are their goals:

- Radical interpreter assigns mental states (including evidence) to human in such a way as to correctly predict human's actions (assuming human is rational).
- Human acts so as to maximise evidential expected utility, where the evidence is what the radical interpreter says the evidence is.

## Key Concepts

- A set of plays is a **Nash equilibrium** if no player can improve their situation by unilaterally changing their play.
- A Nash equilibrium is **Pareto dominant** if no player prefers some other Nash equilibrium to it.
- A Nash equilibrium is **risk dominant** if (roughly) it maximises expected utility assuming other players will randomly choose between equilibrium strategies.