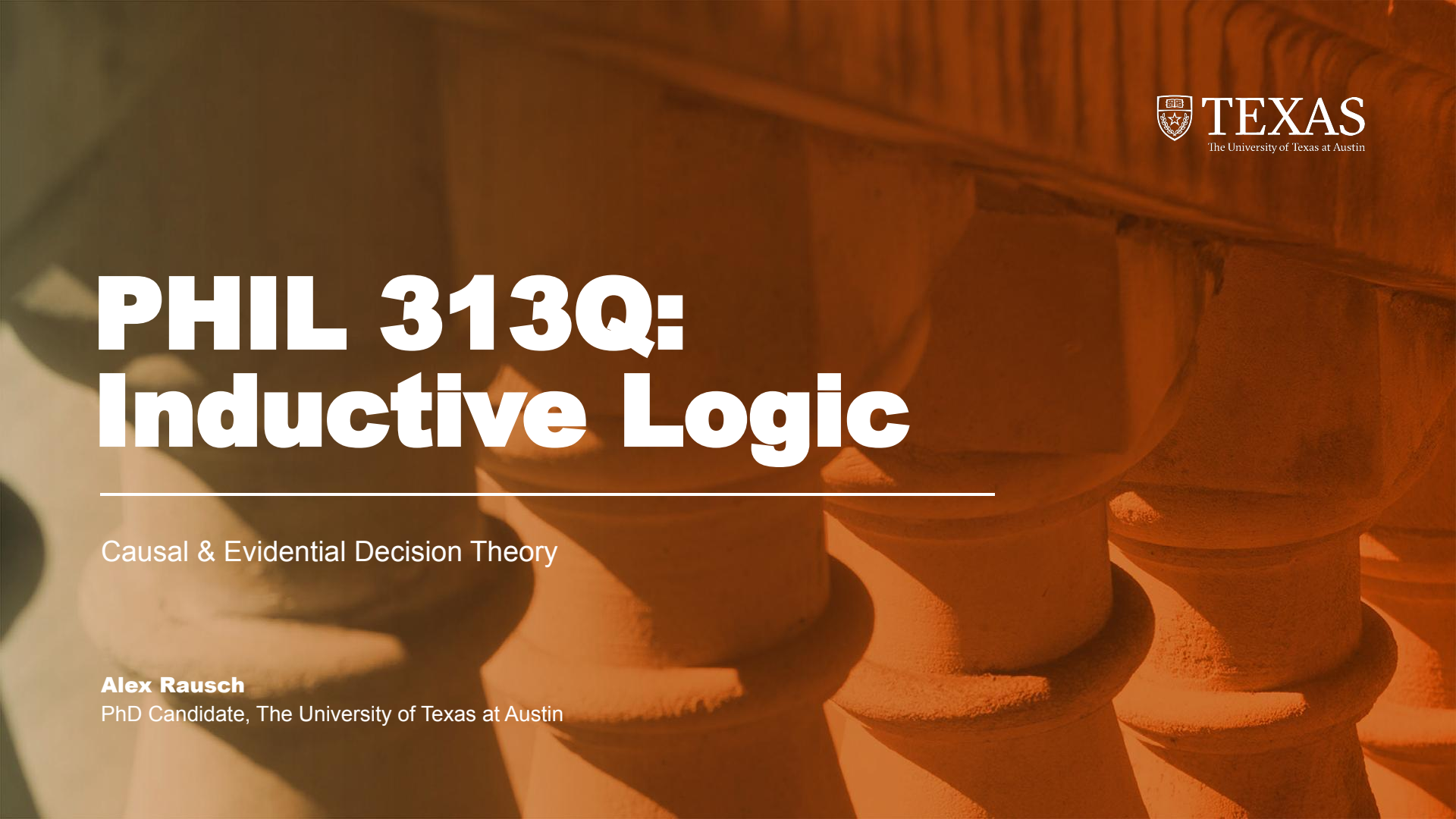

TEXAS
The University of Texas at Austin

# PHIL 313Q: Inductive Logic

Causal & Evidential Decision Theory

**Alex Rausch**
PhD Candidate, The University of Texas at Austin

# Decision Matrices

- I'm traveling in the Netherlands and face a fork in the road. North points to The Hague, where I have no family. East points towards a city ending in "-dam", either Amsterdam or Rotterdam, where I have family in either case.

| | **East is "Amsterdam"** | **East is "Rotterdam"** |
|---|---|---|
| **Go North** | No family<br>-100 | No family<br>-100 |
| **Go East** | Family A<br>+40 | Family B<br>+40 |

- **Go East** "dominates" **Go North** – it has a higher expected value on every partition (in every way the world could be)
- **Dominance Rule**: When deciding between actions, choose whichever one dominates the rest. (If no action dominates the others, calculate expected values.)

# Decision Matrices

- Now imagine that you have a nice family up north (+20), a mean family in Amsterdam (-40), and a super nice family in Rotterdam (+100).

| | **East is "Amsterdam"** | **East is "Rotterdam"** |
|---|---|---|
| **Go North** | Nice family<br>+20 | Nice family<br>+20 |
| **Go East** | Mean family<br>-40 | Super nice family<br>+100 |

- No action dominates the other.
- **Exp(Go North)** = (.5 x 20) + (.5 x 20) = 20
- **Exp(Go East)** = (.5 x -40) + (.5 x 100) = 30
- So, you should go East…if these are the only consequences you care about.

# Pascal's Wager

- Blaise Pascal (17th century) argued that it is rational to act so as to come to believe in a good God.

| | **Good God Exists** | **~ (Good God Exists)** |
|---|---|---|
| **Act so as to believe** | Heaven<br>$\infty$ | Boring life<br>-40 |
| **~(Act so as to believe)** | Hell<br>$-\infty$ | Exciting life<br>+40 |

- **Exp(act so as to believe)** = .5(∞) + (1x-40) = ∞
- **Exp(~ act so as to believe)** = .5(-∞) + (1x40) = -∞

# Pascal's Wager

- The atheist's counter wager:

| | **Bad God Exists** | **~ (Bad God Exists)** |
|---|---|---|
| **Act so as to believe** | Hell<br>$-\infty$ | Boring life<br>-40 |
| **~(Act so as to believe)** | Heaven<br>$\infty$ | Exciting life<br>+40 |

- **Not acting so as to believe** dominates!
- **Exp(act so as to believe)** = .5(-∞) + .5(-40) = -∞
- **Exp(~ act so as to believe)** = .5(∞) + .5(40) = ∞

# Pascal's Wager

- Philosophical issues:
  - Are there separate arguments such as…
    - If God exists, then God is good.
    - If God exists, then God is one.
    - If God exists, then it is the God of tradition X (e.g. Judeo-Christian).
  - Does "acting so as to believe" actually make you believe?
  - Can we change our beliefs at will? (Psychologically? Rationally?)
    - Or do we always need some evidence to change a belief?

# Problems for Our Expected Value Calculations

- 3% of smokers over 65 develop lung cancer
  - **Pr(C|S)** = .03
- .05% of non-smokers over 65 develop lung cancer
  - **Pr(C|~S)** = .005
- Assume smoking is pleasurable (+50) and lung cancer is not (-5,000)
- **Exp(S)** = (1 x 50) + (.03 x -5,000) = -100
- **Exp(~S)** = (1 x 0) + (.005 x -5,000) = -25
- So, don't smoke!
- **But now, suppose it's discovered that a gene causes the desire the smoke and lung cancer *separately*.**
  - Same data as above, but how would you advise me?
  - "It doesn't matter!"

# Problems with Expected Value Calculations

- Here is a decision table:

| | **Have gene** | **~ (Have gene)** |
|---|---|---|
| **Smoke** | Pleasure + High Risk<br>(1 x 40) + (.03 x -5000)<br>-110 | Pleasure + Low Risk<br>(1 x 40) + (.005 x -5000)<br>15 |
| **~(Smoke)** | High Risk<br>(.03 x -5000)<br>-150 | Low Risk<br>(.005 x -5000)<br>-25 |

- **Smoke** dominates!
- What went wrong with the expected value calculation?
  - In this case, smoking doesn't *cause* cancer, it is merely correlated with it!

# Problems with Expected Value Calculations

- **Evidential Decision Theory:** When choose between actions, maximize the expected value of possible consequences *even if they are mere correlations*
  - Problem: the case of the *smoking/cancer gene*
- Perhaps we need to require that, in order for an expected value calculation to "make sense," the consequences must be *caused by* the actions:
- **Causal Decision Theory**: When choosing between actions, maximize expected value of possible consequences *only if they are caused by the actions*
  - As we'll see now, this view also faces some problems.
    - e.g. Prisoner's Dilemma, Newcomb's Paradox

# Prisoner's Dilemma

- Each prisoner wants to maximize his points according to this scheme:

| | **Prisoner B PEACE** | **Prisoner B WAR** |
|---|---|---|
| **Prisoner A PEACE** | A: 5<br>B: 5 | A: -10<br>B: +10 |
| **Prisoner A WAR** | A: 10<br>B: -10 | A: -5<br>B: -5 |

- From both prisoners' perspective, choosing WAR dominates. But *if they both reason this way*, they continually lose points.
- The prisoners' choices don't *cause* one another, so Causal Decision theory entails no expected value calculations. But if the prisoners reason similarly (90% of the time):
  - **Exp(Peace)** = (.9 x 5) + (.1 x -10) = 3.5
  - **Exp(War)** = (.9 x -5) + (.1 x 10) = -4.5 + 1 = -3.5
- So Evidential Decision theory predicts that each should PEACE (which is correct?)

# Newcomb's Paradox

**Box A**


**Box B**


- You will need to decide between two options.
  - **TWO BOX:** Take Box B *and* Box A
  - **ONE BOX:** Take Box B
- The catch: before you enter the room, we perform a brain scan and predict with a high degree of confidence (99%) what you'll do.
  - **If we predict you TWO BOX, we put NOTHING in Box B before you enter.**
  - **If we predict you ONE BOX, we put $1 MILLION in Box B before you enter.**

# Newcomb's Paradox

**Box A**


**Box B**


- At this point, many people choose **ONE BOX**.
- Expected value calculations agree with this intuition:
  - **Exp(ONE BOX)** = (.99 x 1,000,000) + (.01 x 0) = $990,000
  - **Exp(TWO BOX)** = (.99 x 1,000) + (.01 x 1,001,000) = $11,000
- But notice: **your action itself has no cause on the consequences (the payouts)**
  - What is in Box B is determined prior to your action, so can't be caused by it.
- So while Evidential Decision Theory tells you to ONE BOX, Causal Decision Theory will tell you *no expected value calculations are legitimate*.

# Newcomb's Paradox

**Box A**


**Box B**


- In fact, suppose that the brain scan is complete and you enter the room. The Causal Decision Theorist makes this argument to you:

| | **Predicted ONE BOX** | **Predicted TWO BOX** |
|---|---|---|
| **You ONE BOX** | $1,000,000 | $0 |
| **You TWO BOX** | $1,001,000 | $1,000 |

- **TWO BOX** dominates! *Why wouldn't you take both boxes at this point?*

## Back to Basics

- We've learned about how to determine probabilities, categorical and conditional.
- Before any flips: **"The fair coin has a 50% probability of landing heads."**
- This seems to be a claim *about the coin itself*, and so, *about the world*.
  - We might explain it by appealing to the results of multiple, theoretical flips "in the long run".
- But what about: "**The fair coin has a 50% probability of landing heads *today at noon.*"**
  - Still a claim *about the coin*. We might explain it by appealing to the laws of physics and the physical properties of the setup and coin itself.
- But we also say things like: **"The probability that the dinosaurs became extinct due to an asteroid is 90%."**
  - This is a claim about our degree of confidence in a theory given available evidence.

## Back to Basics

- **Objective Probability:** chance "out in the world," independent of our degrees of belief/confidence
- **Subjective Probability:** chance "due to ignorance," dependent on our degrees of belief

**Philosophical Questions**

1. Does Objective Probability make any sense?
2. Do we live in a "chance"-y world, or is all chance a feature of our ignorance?
3. What do you *mean* when you talk about the probability of something?
    a. Are you reporting on a feature of the world, or on your own degree of confidence?
        i. If the latter, do you *ever* use "probability"-talk to make a claim about the world and *not just* your degree of confidence?