# What I learned in ethics class

The class was a “for adult professionals” version of:
https://web.stanford.edu/class/cs182/

# Hello!

## *I am Lexi Galantino*

I am not an ethics expert!
I took one ethics course. But here, I'll tell you about it.

You can find me at @gallexi

# Ground Rules

1

# Stay Curious

Even if someone's view is very different from yours, you may learn something new about why they feel so differently.

Thank you Lara Hogan for this ground rule!

2

# Remember the real world

For some, some topics posed here seem like interesting thought experiments. For others, they are life-or-death questions with extremely real consequences.

3

# Take a break if you need one

We're going to be discussing topics either directly involving or adjacent to topics that are really vitally important to a lot of people. If you find yourself needing a break, please take one.

4

# Keep this confidential to GH

Keeping thoughts shared here to our in-group helps us work through these issues, I think.

# Agenda

- Story: Omelas
- Multiple ethical models
- Types of fairness
  - Discussion
- Silicon Valley <-> The Defense Sector
  - Discussion
- Story: The Ones Who Stay and Fight
- General discussion

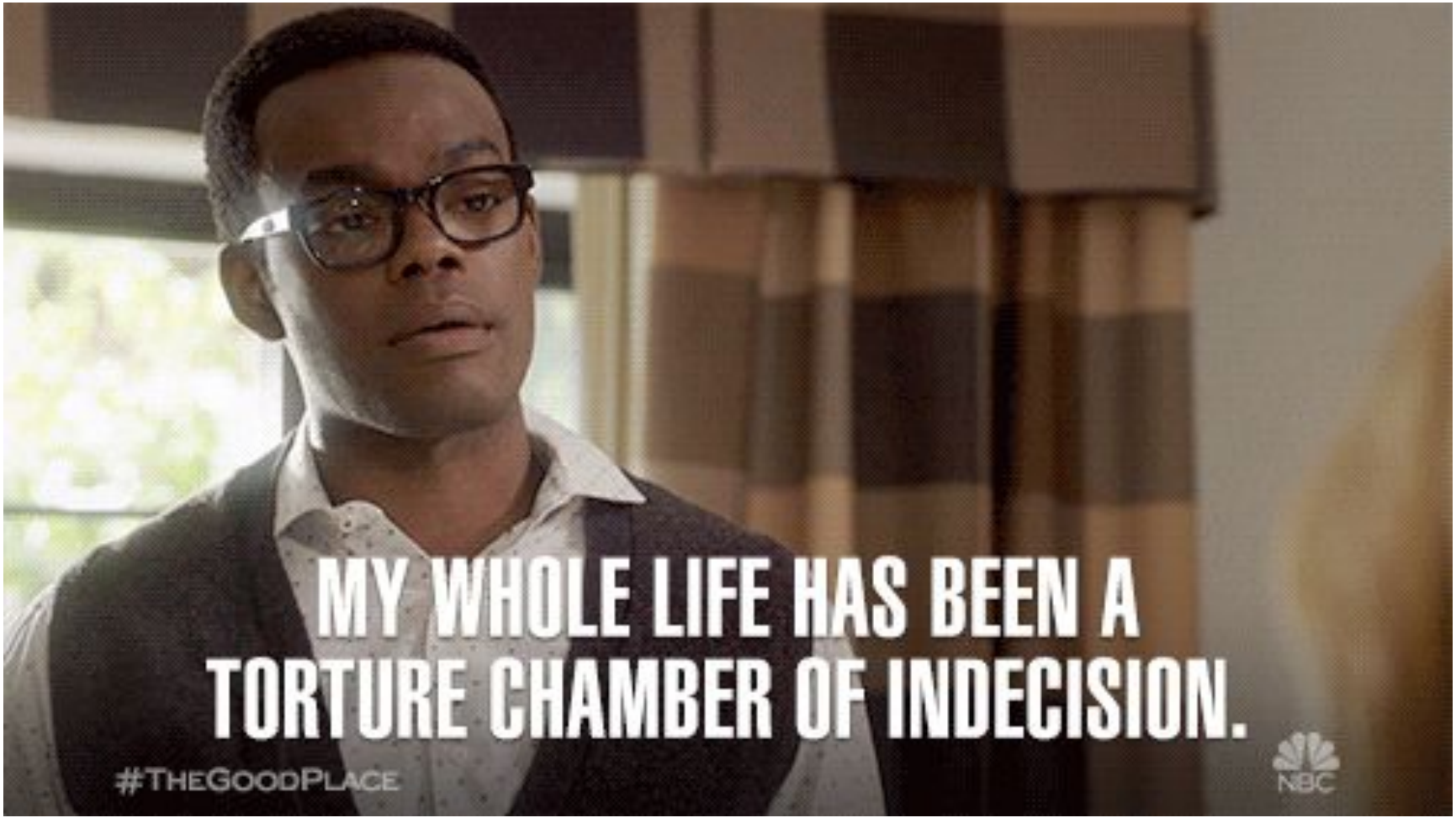
# The Ones Who Walk Away from Omelas

by Ursula K. Le Guin

Absolute paradise for almost every single person, save for one child living in misery for their entire life. If you rescue the child, everything else comes crashing down for everyone. Most people accept this; some leave.

# No single “ethical” choice

In fact, there are multiple ways to determine right from wrong, and humans haven’t decided that one is the “more ethical” one


MY WHOLE LIFE HAS BEEN A
TORTURE CHAMBER OF INDECISION.
#THEGOODPLACE
NBC

# Two major systems

## Deontology

"an ethical theory that uses rules to distinguish right from wrong"

https://ethicsunwrapped.utexas.edu/glossary/deontology

## Consequentialism

"an ethical theory that judges whether or not something is right by what its consequences are"

https://ethicsunwrapped.utexas.edu/glossary/consequentialism

# Two subsets

## Moral absolutism

Subset of deontology

“Lying is wrong because it’s wrong”

https://ethicsunwrapped.utexas.edu/glossary/moral_absolutism

## Utilitarianism

Subset of consequentialism

“Lying is ok if it does good”

https://ethicsunwrapped.utexas.edu/glossary/utilitarianism


Okay, I've made my decision.
I want to, start crying.

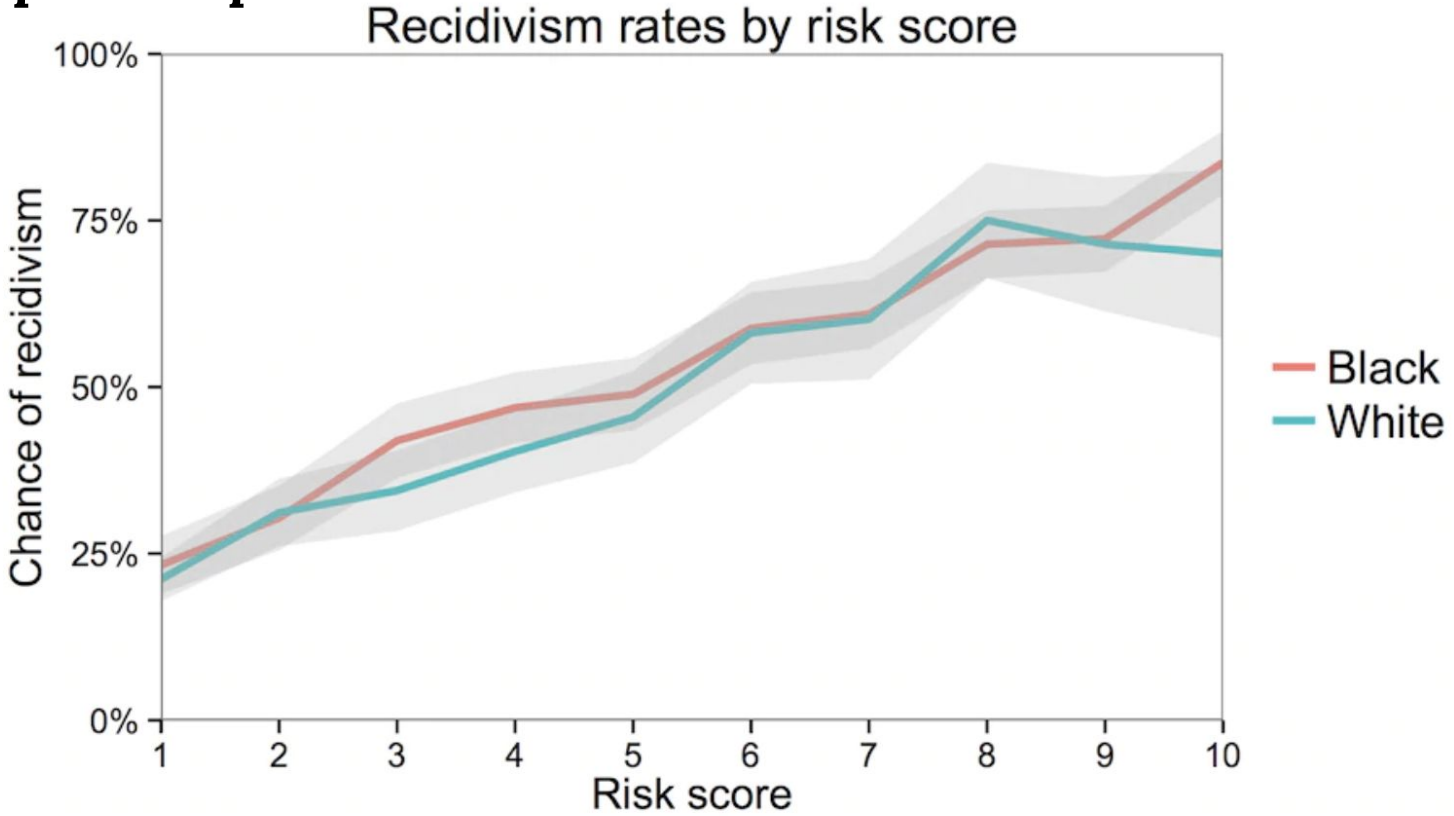
# Fairness

There is also no one way to define fairness, at least mathematically. Instead, there are at least 21. Let's talk about two

## Case study: COMPAS

- Algorithm designed to predict recidivism
- ProPublica found it was biased against Black people
- Northpointe says that it is not biased
- Each group is using a different definition of fairness

**Northpointe says it's fair because it has the same accuracy regardless of defendant race, the same "positive predictive value"**



# But, ProPublica found that when it fails, it fails differently for Black people than for white people

| | Black people | White people |
|---|---|---|
| False Positive | 44.9% | 23.5% |
| False Negative | 28% | 47.7% |

# Two definitions of fairness

## Calibration

"Outcomes should be independent of protected characteristics conditional on risk scores"

The model has the same accuracy for both white and Black defendants: 61%

$Pr(Y = 1 | s(X), Xp) = Pr(Y = 1 | s(X))$

## Classification Parity

"Classification error is equivalent across groups defined by protected characteristics"

The model *does not have* the same false positive rate across race groups

$Pr(d(X) = 1 | Y = 0, Xp) = Pr(d(X) = 1 | Y = 0)$

**So, they should just rework the algorithm so it has both!**

**Big problem?**

**Kleinberg et al (2017) showed that you can't (generally) have both.**

# Discussion

- Under what conditions would you feel comfortable with an algorithm making a critical decision like this in your life?

# Ethics of Defense Tech

# The Ethics of Defense Technology Development: An Investor's Perspective

Trae Stephens Follow
Dec 4, 2019 · 14 min read ★

## Key Argument 1: We're better than the alternative

- If the US isn't a dominant power, including with top defense tech, that leaves a vacuum for Russia and China to fill, and they will do so

- We're the only one of the three that is a democracy

- "American technologists are not required to work on behalf of their nation's defense, but in choosing not to do so, they must recognize that they are ceding an advantage to illiberal rivals and putting the very freedom and openness that they cherish at risk."

## Key Argument 2:
## The Just War Theory

- Just because using code in warfare is new doesn't mean humans thinking about ethics in war and defense is
- Just War Theory is thousands of years old, has been debated by generations of humans in nearly every faith and tradition
- Principles of Just War Theory have informed the US Law of War, the Geneva Conventions, the UN Charter, and more

## The Principle of Last Resort

- Loss of life should be avoided if at all possible
- Tech can advance defensive technologies like early warning systems that significantly increase the cost of "trying anything"
- There are also advanced non-lethal and "non-kinetic" weapons (like cyber-warfare) that can further help avoid lethal violence

*“Technology development will play a critical deterrent role in a post-nuclear society, both as a neutralizer of nuclear and conventional threats as well as in shifting the cost of conflict to the aggressor.”*

- Trae Stephens

## The Principles of Discrimination and Proportionality

- When force is required, it should be deployed according to principles of who is a legitimate target and how much force is morally defensible to use
- Weapons have historically gone:
  - one to one – swords
  - one to some – cannons
  - one to many – nuclear
- Tech can help us come back to one to some or some to some, with more precision: drones, battlefield awareness

*“Technology has the potential to lead to a significant reduction in the loss of innocent life through highly precise (discriminating) and targeted attacks and for these surgical operations to reduce the need for massive indiscriminate (disproportionate) strikes.”*

- Trae Stephens

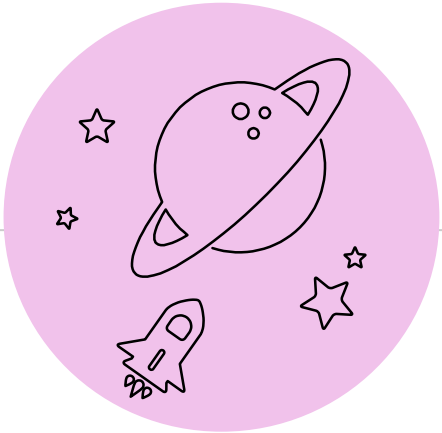
## The Principles of Just Authority and Right Intent

- Critical idea is that it is the trust of the people that give government the right to hand down “punishment”
- With better tech, we can have more accurate intelligence, and counter misinformation
- “The principle of Just Authority depends critically on thoughtful policy and open communication around the development and use of defense technology.”
- “It is therefore critical, ...to lay in place clear policies and practices for the ways in which these technologies can be morally and justly used.”

# Discussion

- “Many in Silicon Valley hold the mistaken belief that if they and their counterparts withdraw from defense or weapons work, they can force a stoppage and bring about a peaceful equilibrium. There is a fundamental consideration that has been too little covered in this debate, however: What are the moral consequences of societies rooted in a Just War tradition refusing to invest in sophisticated defense technologies while authoritarian regimes invest aggressively in their development?”

- What are your thoughts?



By N.K. Jemison

# The Ones Who Stay And Fight

Also paradise, but this one is super intentional about diversity, not just tolerance. The big downside is that anyone who discovers the hateful ideas that were common in the past (our time) must be quickly and absolutely killed. A father is killed in this way in front of his child, and the child explodes in anger and pain at the “social workers”, vowing vengeance. They take her under their wing to become one of them so she doesn’t have to be killed too.

*“So don’t walk away. The child needs you, too, don’t you see? You also have to fight for her, now that you know she exists, or walking away is meaningless.”*

– N.K. Jemison

# Discussion

- What is something that surprised you?
- Is there a view that is different from your own that you are better able to understand?
- Is there a view that you hold that you can put in context and terminology of this discussion?

## Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival