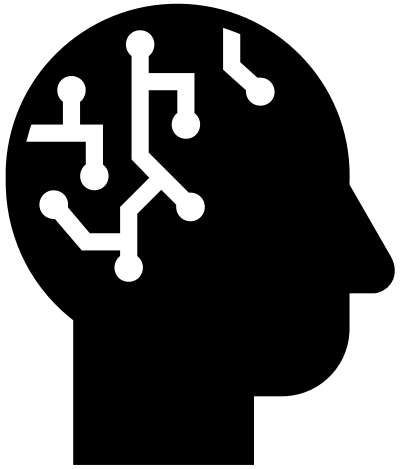
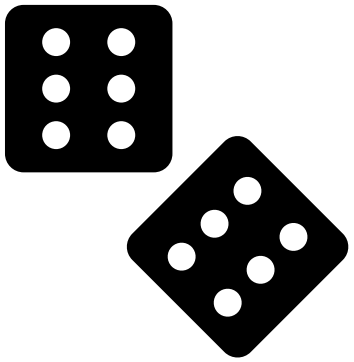



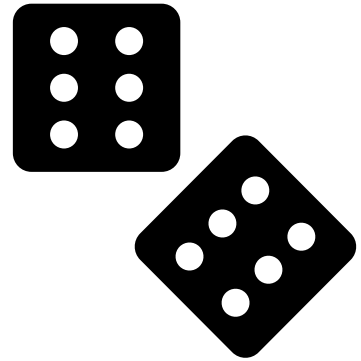

Ethics: Probability and AI

Dr. Justin Shin

McCoy Family Center for Ethics in Society

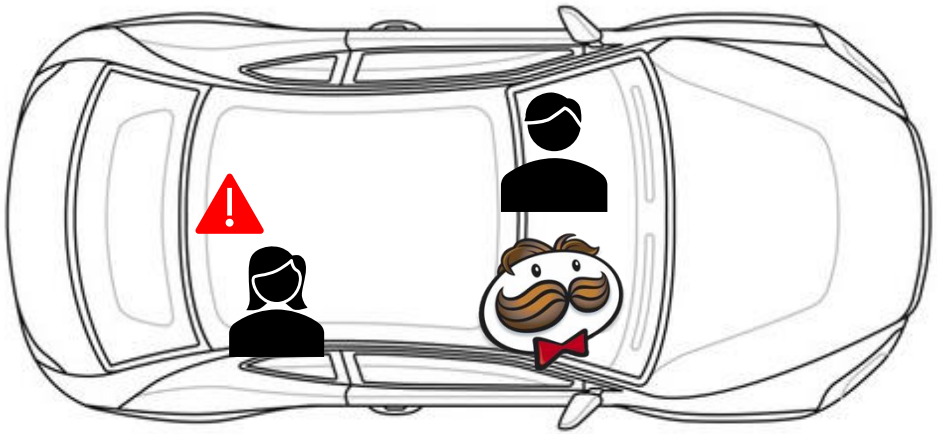
Stanford University

Themes for today…

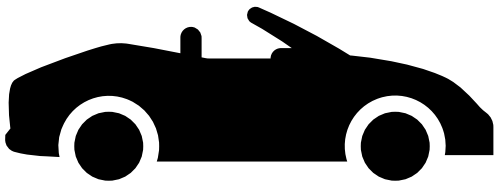


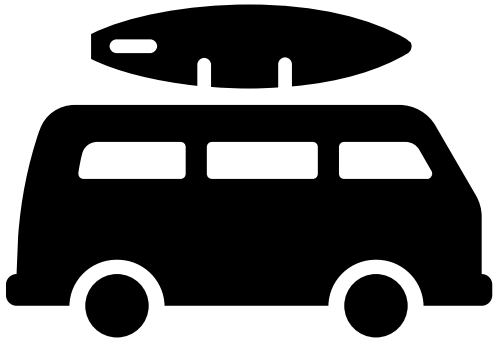
- How do probabilities show up in ethics problems?
- What, if anything, makes probabilistic evidence different from other forms of evidence?
- This course addresses some of the probabilistic foundations of AI. What sorts of concerns does AI inherit from these probabilistic foundations?

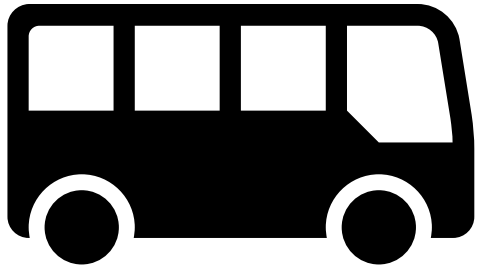
Maryland v. Pringle



How Should We Interpret the Probable of Probable Cause?







A Theoretical Aside… Blackstone's Ratio

"…all presumptive evidence of felony should be admitted cautiously, for the law holds that it is better that ten guilty persons escape than that one innocent suffer."

Why the Asymmetry?

Why is a guilty person going free bad?

The person who did the crime goes unpunished.

Why is an innocent person being convicted bad?

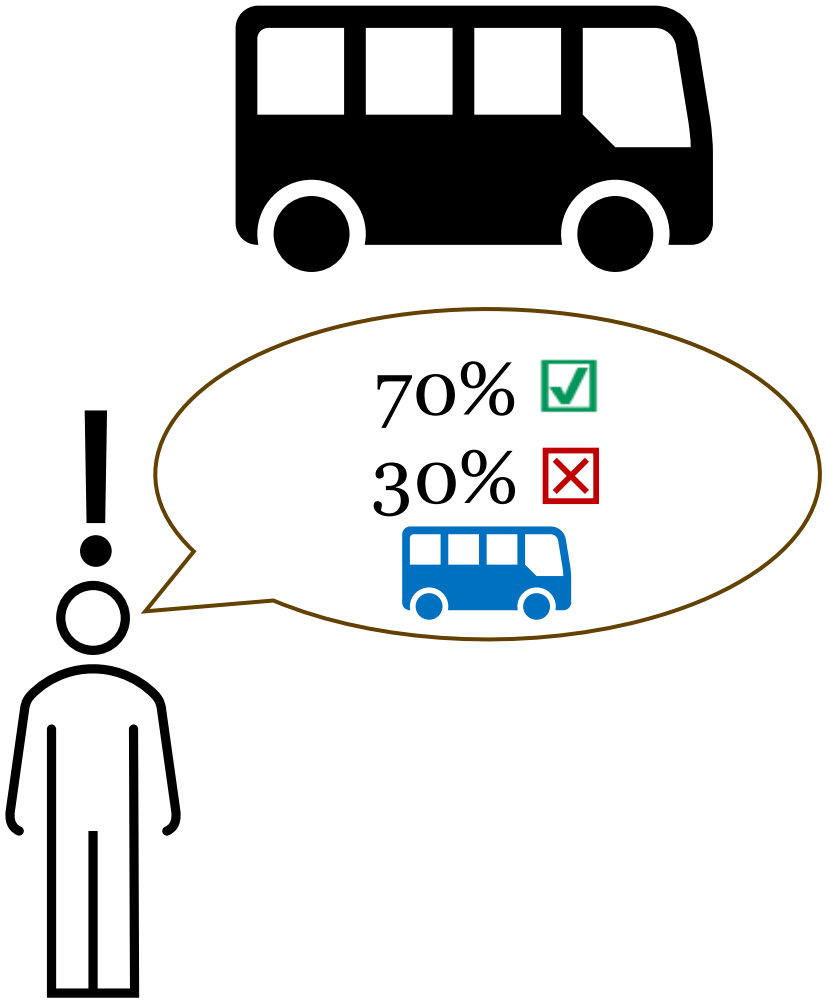
The person who did the crime goes unpunished.

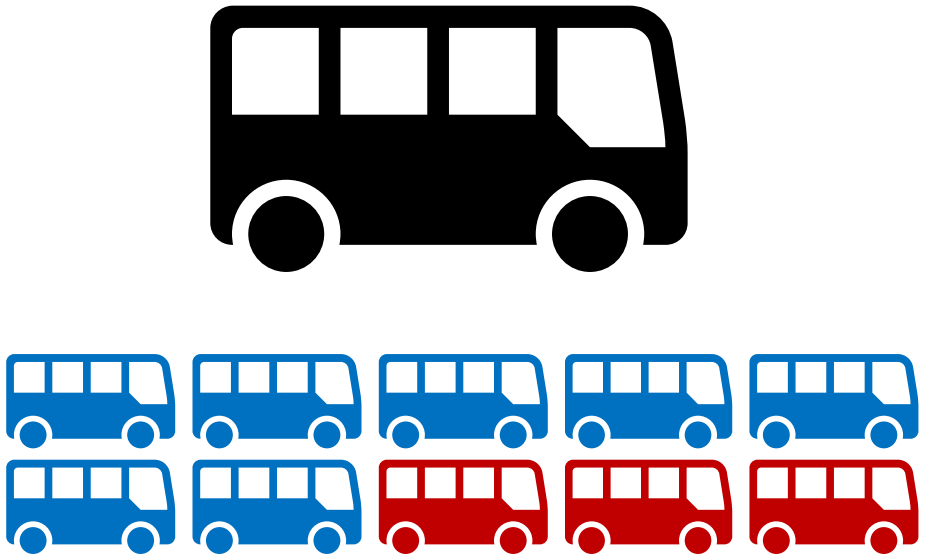
And also, an innocent person is punished undeservedly.

Isn’t this just a problem of vagueness?

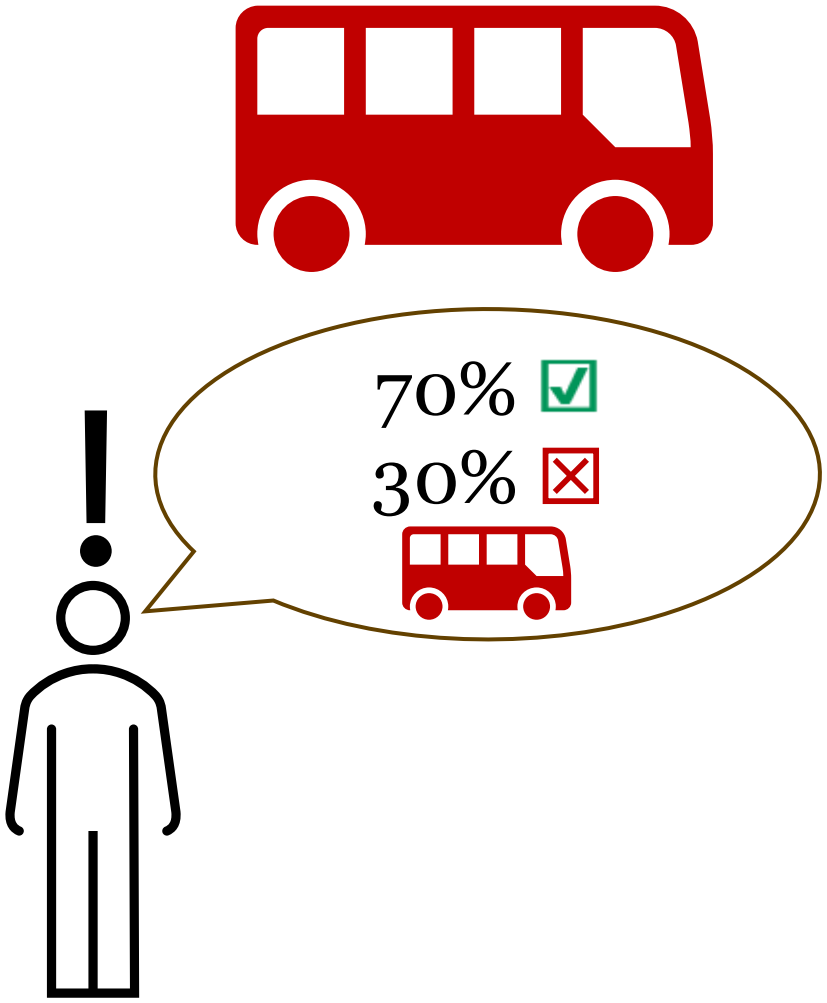
Is there any issue of statistical evidence that arises if we grant precise probabilities?

Probabilistic Evidence (Smith v. Rapid)





Probabilistic Evidence (Smith v. Rapid)



Individualization Gets Messy

#	PREDICTOR RAT VICTIM SHOOTING INCIDENTS	Attribute RAW Score; The raw score calculated by the model; The number of times an individual has been the victim of a shooting.
#	PREDICTOR RAT VICTIM BATTERY OR ASSAULT	Attribute RAW Score; The raw score calculated by the model; The number of times an individual has been the victim of an aggravated battery or aggravated assault; .

34. Was your mother (or mother figure who principally raised you) ever arrested, that you know of?
☑ No ☐ Yes

56. Do you have a regular living situation (an address where you usually stay and can be reached)?
☐ No ☑ Yes

Individualization Gets Messy

Rideshare drivers sue Uber over being kicked off app in new challenge to California law

BY LEVI SUMAGAYSAY

APRIL 20, 2026 UPDATED APRIL 22, 2026

Republish

Do These Sorts of Models Encourage Punishing Who You Are Rather Than What You Do?

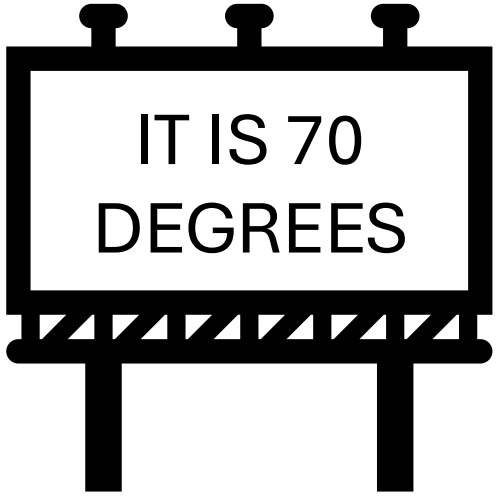
Robinson v. California

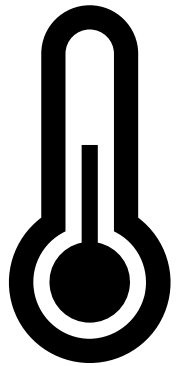
California Health and Safety Code § 11721: No person shall use, or be under the influence of, or be addicted to the use of narcotics, excepting when administered by or under the direction of a person licensed by the State to prescribe and administer narcotics.

“The present case, Robinson versus California, is of challenge to and attack upon Section 11721 of the Health and Safety Code of the State of California essentially because we feel that it is a denial of equal protection and due process in that it punishes a status rather than an act or omission, that it punishes an involuntary status, that it punishes a status of physical and mental illness… it is ex post facto and that it imposes cruel and unusual punishment.”

Commonwealth v. Canadyan

Probabilistic Systems and Epistemic Luck





Probabilistic Systems and Epistemic Luck

- **Top-p:** Limits the pool of potential words to a subset that cumulatively meets a certain probability threshold.
- **Top-k:** Caps the word choices to a fixed number of the highest-probability candidates.

60 degrees 15%
61 degrees 13%
59 degrees 8%
62 degrees 6%
…

Probabilistic Systems and Epistemic Luck

AI Hallucinates Flight Refund, Sparking Memes in China

A chat log in which an AI assistant confidently presented a user with numerous inaccuracies has inspired a wave of AI hallucination spoofs, a growing trend in the country.

Update on the ChatGPT Case: Counsel Who Submitted Fake Cases Are Sanctioned

Air Canada Has to Honor a Refund Policy Its Chatbot Made Up

The airline tried to argue that it shouldn't be liable for anything its chatbot says.

Some more basic worries about probabilities...

A jury found defendant Malcolm Ricardo Collins and his wife defendant Janet Louise Collins guilty of second degree *321 robbery (Pen. Code, §§211, 211a, 1157). Malcolm appeals from the judgment of conviction. Janet has not appealed.[1]

On June 18, 1964, about 11:30 a.m. Mrs. Juanita Brooks, who had been shopping, was walking home along an alley in the San Pedro area of the City of Los Angeles. She was pulling behind her a wicker basket carryall containing groceries and had her purse on top of the packages. She was using a cane. As she stooped down to pick up an empty carton, she was suddenly pushed to the ground by a person whom she neither saw nor heard approach. She was stunned by the fall and felt some pain. She managed to look up and saw a young woman running from the scene. According to Mrs. Brooks the latter appeared to weigh about 145 pounds, was wearing 11 something dark," and had hair "between a dark blond and a light blond," but lighter than the color of defendant Janet Collins' hair as it appeared at trial. Immediately after the incident, Mrs. Brooks discovered that her purse, containing between $35 and $40 was missing.

Some more basic worries about probabilities…

Black man with beard	1 in 10
Man with mustache	1 in 4
White woman with pony tail	1 in 10
White woman with blond hair	1 in 3
Yellow motor car	1 in 10
Interracial couple in car	1 in 1,000

Some more basic worries about probabilities…


I didn't kill my babies

In November 1999 solicitor Sally Clark, 36, was convicted of one of the most horrific crimes any mother could imagine – the murder of her baby sons, Christopher, aged 11 weeks, and Harry, eight weeks.

When he died, a part of me died too

1. The ratio of live births to cot deaths is 8,500:1, and so, the probability that a random infant dies of cot death is about 1/8,500.
2. Therefore, if a family has two infants, the probability that both will die from cot death is $(1/8,500)^2$, or about 1/72,000,000.
3. The birth rate of Great Britain is about 800,000 a year, and so we should only expect about 5 genuine incidents of double cot deaths every 460 years.
   (800,000/72,000,000)*450=5
4. The rate of double homicides in Great Britain is much greater than 5 in 450 years, and so it is more likely that Sally Clark murdered her two children!

Some Conclusions…

Probabilistic interpretations of principles of justice and individual rights are difficult to pin down. (Maryland v Pringle)

Probabilistic evidence faces difficulties in meeting individualization requirements.

Some probabilistic models in law approach the issue of punishing a status as opposed to an act. Careful selection of relevant inputs is needed to avoid this. The test of counterfactuals applied may be relevant.

The probabilistic nature of some AI models introduce some challenges to do with epistemic luck. Hallucinations of standard models can be attributed to their probabilistic nature, and signal epistemic unreliability.

A bonus issue: The Prosecutor’s Fallacy.

https://tinyurl.com/embeddedethicsS26

10-15 minute survey, taking it (or not) won't impact your grade in the class in any way. The teaching team won't know who participates.

Option to provide your **Stanford email address** to receive a **$10 gift card**, **up to the first 800 participants**. Compensation once per quarter (SUNet login required).

Questions? Email embeddedethics@stanford.edu

Stanford | Embedded Ethics