Common Pitfalls for Studying the Human Side of Machine Learning

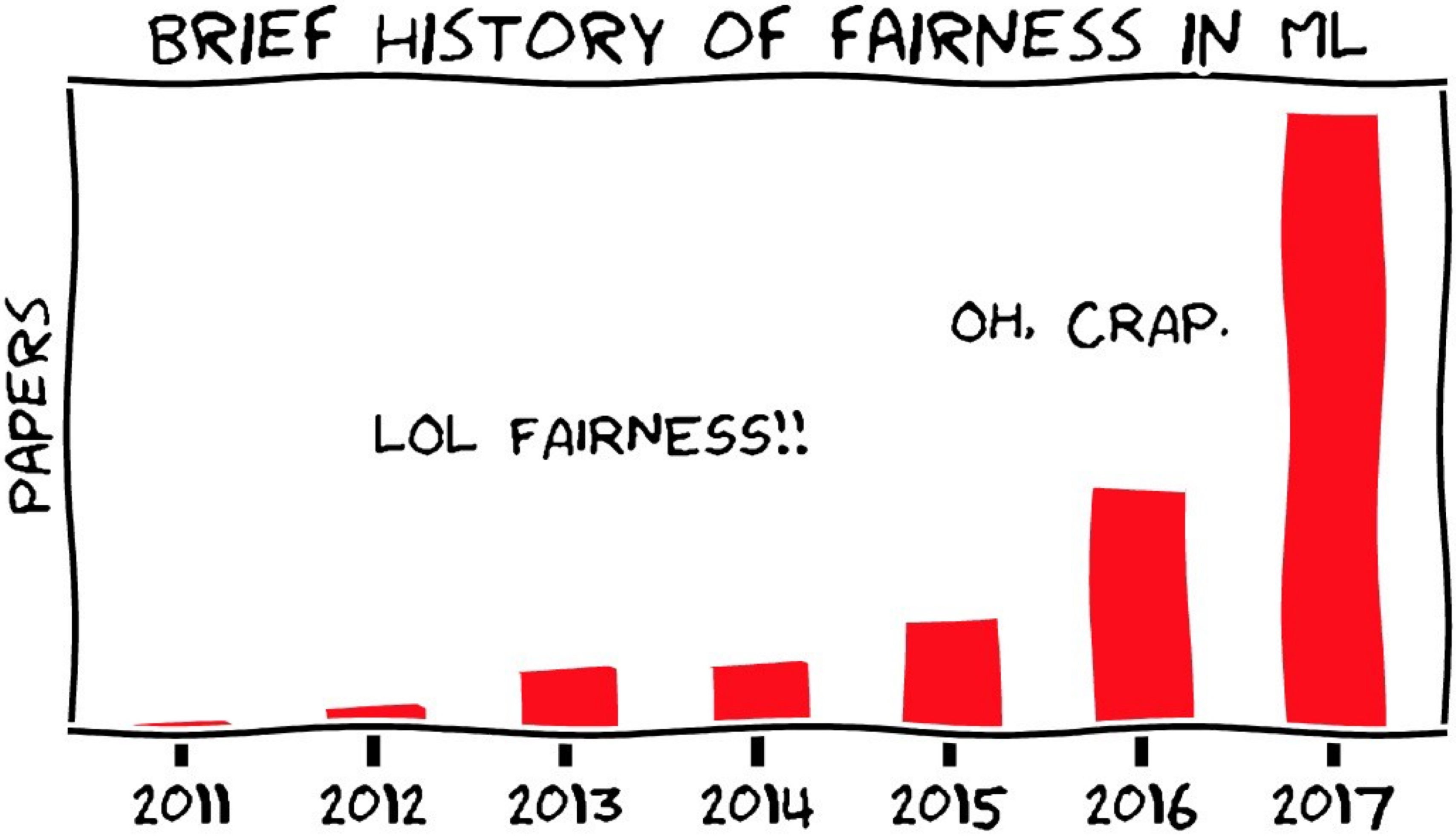
Joshua A. Kroll, **Nitin Kohli**, Deirdre Mulligan

UC Berkeley School of Information

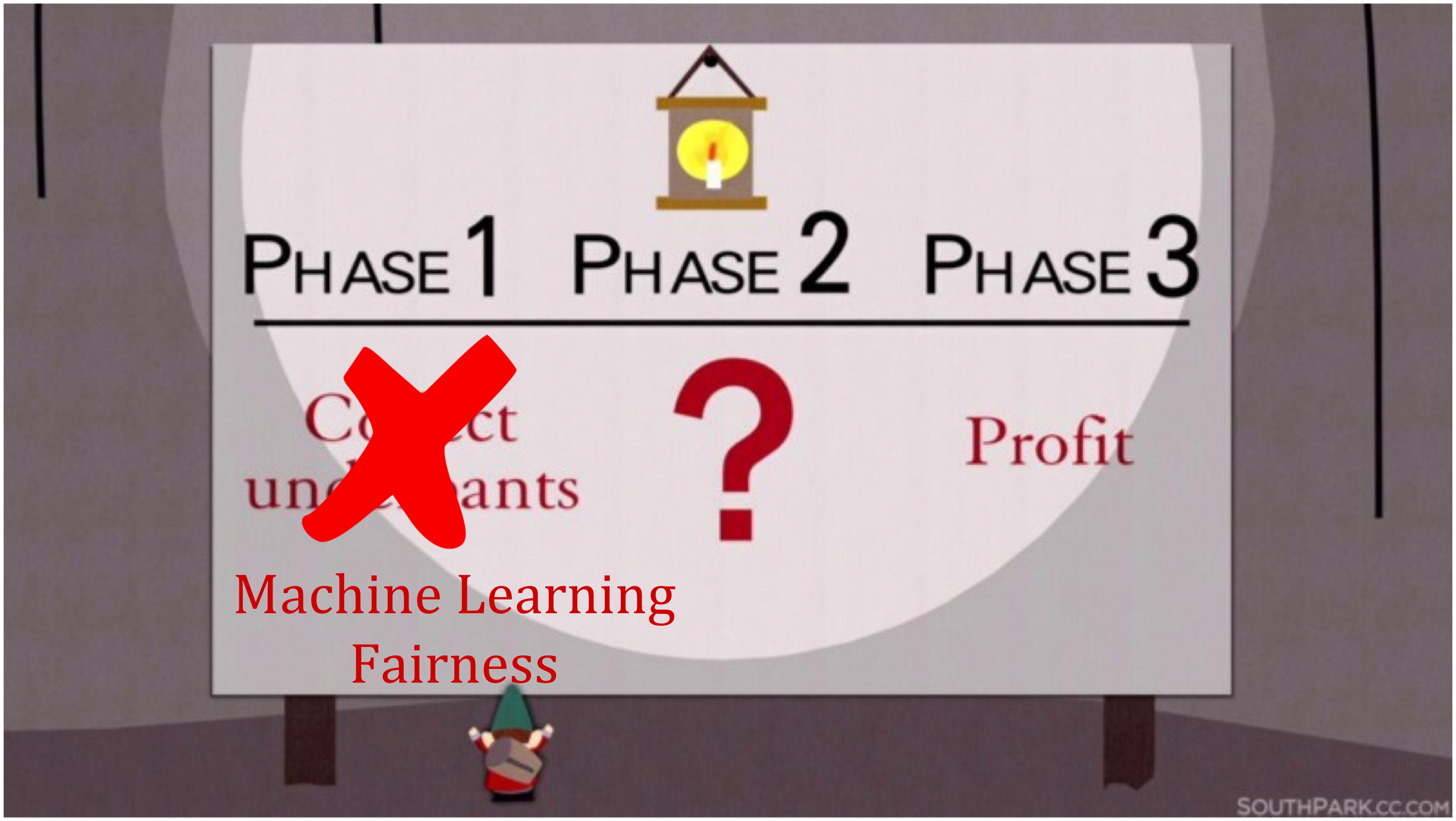
Tutorial: NeurIPS 2018

3 December 2018

Berkeley SCHOOL OF INFORMATION



Credit: Last Year, Solon Barocas and Moritz Hardt, "Fairness in Machine Learning", NeurIPS 2017


PHASE 1
PHASE 2
PHASE 3
~~Collect underpants~~
?
Profit
Machine Learning Fairness
SOUTHPARK.CC.COM

What goes wrong when engaging other disciplines?

- Want to build technology people can *trust* and which supports *human values*
- Demand for:
  - ***Fairness***
  - ***Accountability***
  - ***Transparency***
  - ***Interpretability***
- These are rich concepts, with long histories, studied in many ways

- But these terms get re-used to mean different things!
  - This causes unnecessary misunderstanding and argument.
  - We'll examine different ideas referenced by the same words, and examine some concrete cases

Why this isn’t ethics

Machine learning is a tool that solves specific problems

Many concerns about computer systems arise not from people being unethical, but rather from misusing machine learning in a way that clouds the problem at hand

Discussions of ethics put the focus on the individual actors, sidestepping social, political, and organizational dynamics and incentives

Definitions are unhelpful
(but you still need them)

Values Resist Definition

Definitions aren’t for everyone:
Where you sit is where you stand

If we’re trying to capture human values,
perhaps mathematical correctness isn’t enough

These problems are *sociotechnical* problems

Fairness

“What is the problem to which fair machine learning is the solution?” - Solon Barocas

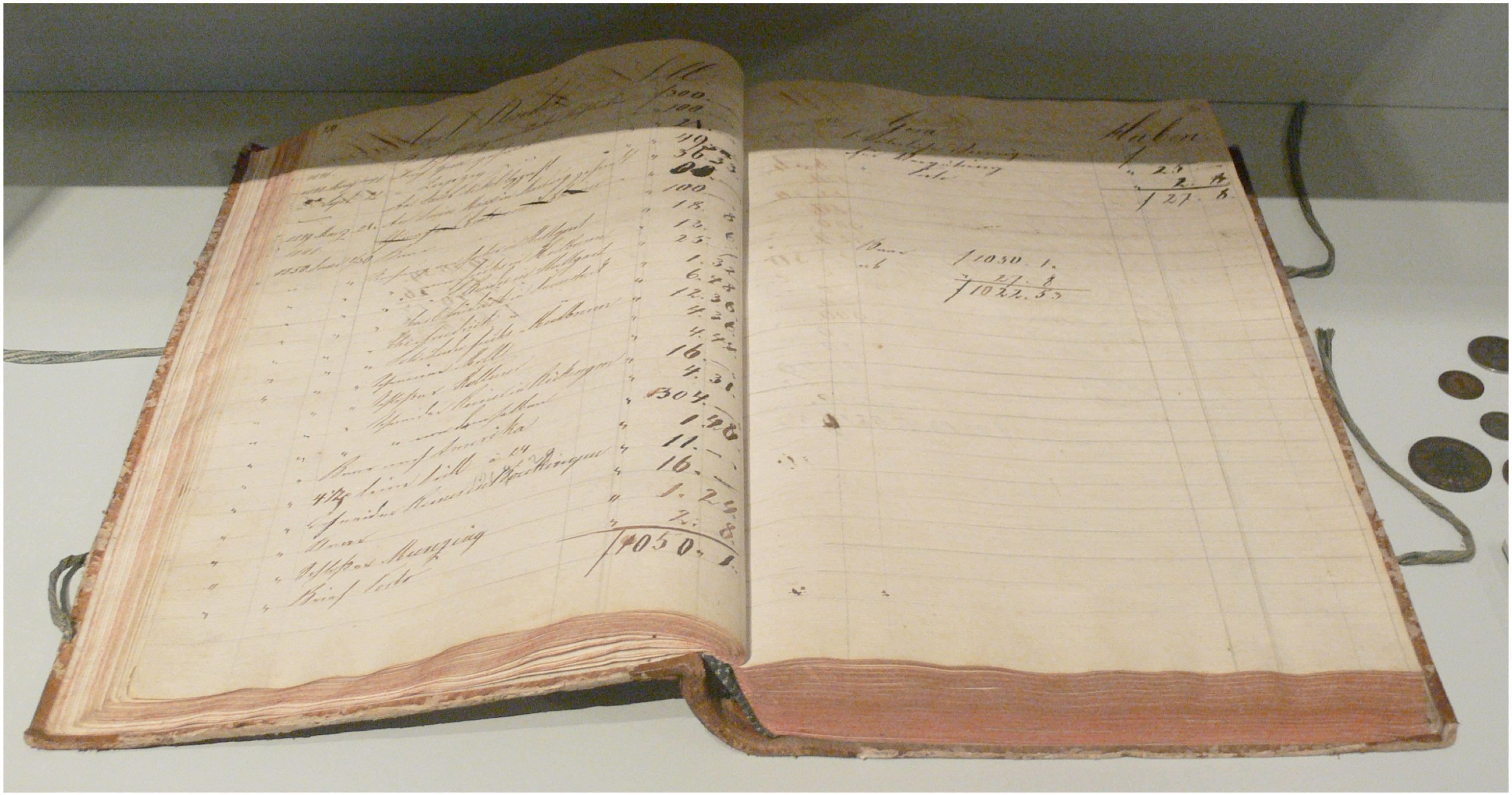
What is Fairness:
Rules are not processes

Tradeoffs are inevitable

Maybe the Problem is Elsewhere

What is Accountability: Understanding the Unit of Analysis

What should be true of a system, and where should we intervene on that system to guarantee this?



Haben

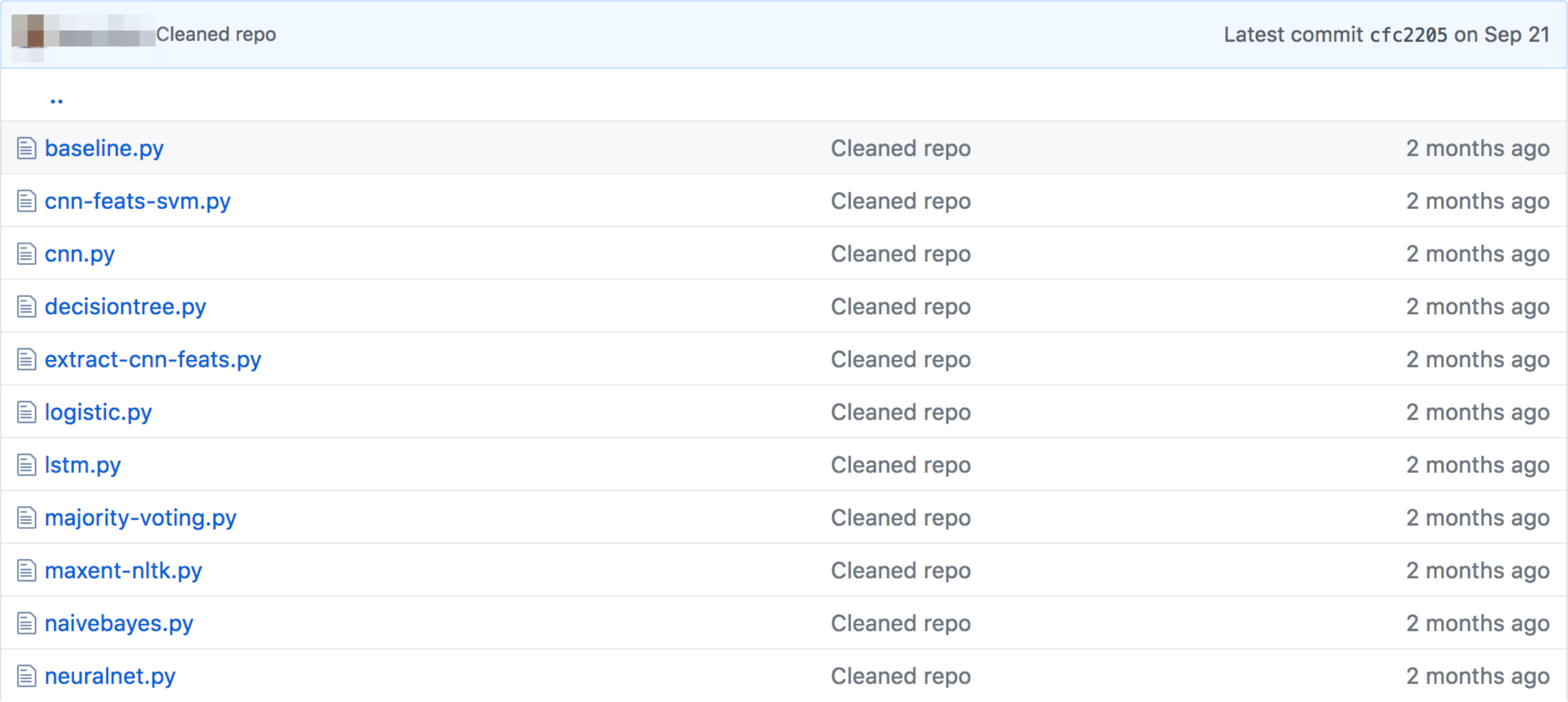
	500.	
	100.	
	40	
	36.	33
	100.	
	18.	
	13.	6
	25.	
	1.	34
	6.	48
	12.	36
	4.	36
	4.	44
	16.	
	4.	31
	304.	
	1.	48
Amerika	11.	–
	16.	–
	1.	24
	7.	8
	f1050.	1.

	25.	
	2.	4
	f27.	8

	f1050.	1.
	27.	8
	f1022.	53

Transparency & Explainability are Incomplete Solutions

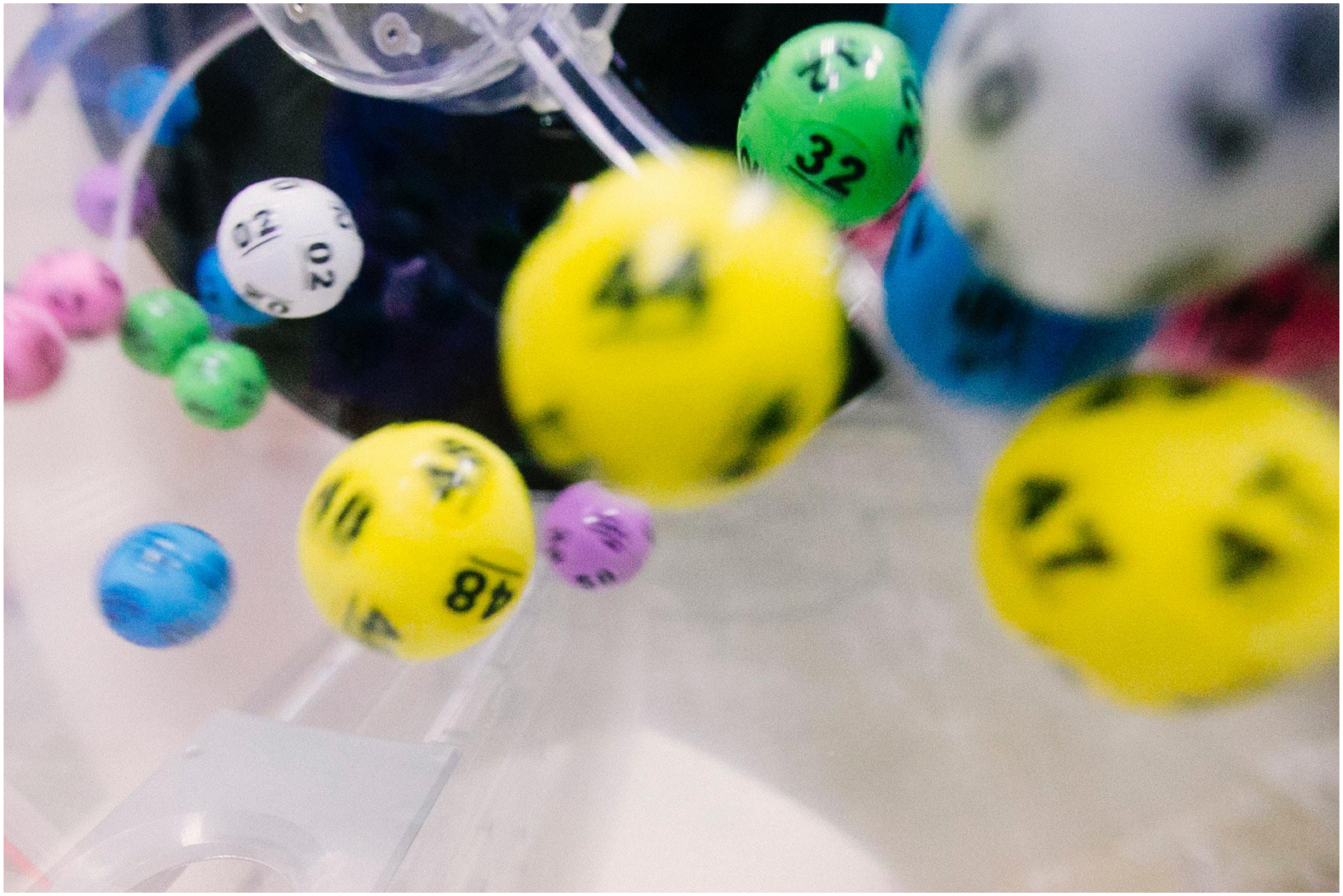
Transparency



[redacted] Cleaned repo — Latest commit `cfc2205` on Sep 21

..

baseline.py	Cleaned repo	2 months ago
cnn-feats-svm.py	Cleaned repo	2 months ago
cnn.py	Cleaned repo	2 months ago
decisiontree.py	Cleaned repo	2 months ago
extract-cnn-feats.py	Cleaned repo	2 months ago
logistic.py	Cleaned repo	2 months ago
lstm.py	Cleaned repo	2 months ago
majority-voting.py	Cleaned repo	2 months ago
maxent-nltk.py	Cleaned repo	2 months ago
naivebayes.py	Cleaned repo	2 months ago
neuralnet.py	Cleaned repo	2 months ago


32
02
44
48
47

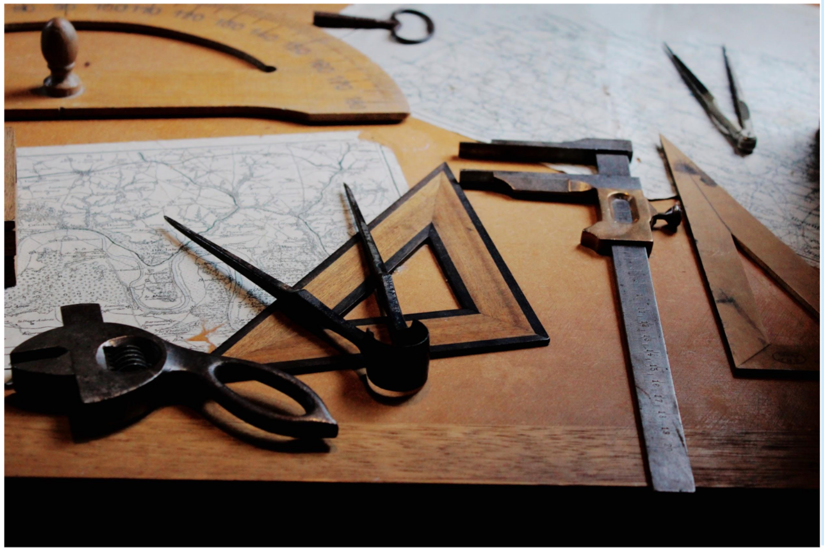
Explainability

Explanations from Miller (2017)

- Causal
- Contrastive
- Selective
- Social
- Both a product and a process

Miller, Tim. "Explanation in artificial intelligence: Insights from the social sciences." arXiv preprint arXiv:1706.07269 (2017).

Data are not the truth



FRANCE
Barcelona
MADRID
PORTUGAL
Palma
Murcia
Sevilla
Gibraltar
Ceuta
Melilla
ALGERIA
MOROCCO
www.comersis.com

If length is hard to measure,
what about unobservable
constructs like risk?

Construct Validity

Abstraction is a fiction

There is no substitute for solving the problem

You must first understand
the problem

Case One :
Babysitter Risk Rating

Xcorp launches a new service that uses social media data to predict whether a babysitter candidate is likely to abuse drugs or exhibit other undesirable tendencies (e.g. aggressiveness, disrespectfulness, etc.)

Using computational techniques, Xcorp will produce a score to rate the riskiness of the candidates. Candidates must opt in to being scored when asked by a potential employer.

This product produces a rating of the quality of the babysitter candidate from 1-5 and displays this to the hiring parent.

With a partner, examine the validity of this approach. Why might this tool concern people, and who might be concerned by it?

What would it mean for this system to be fair?

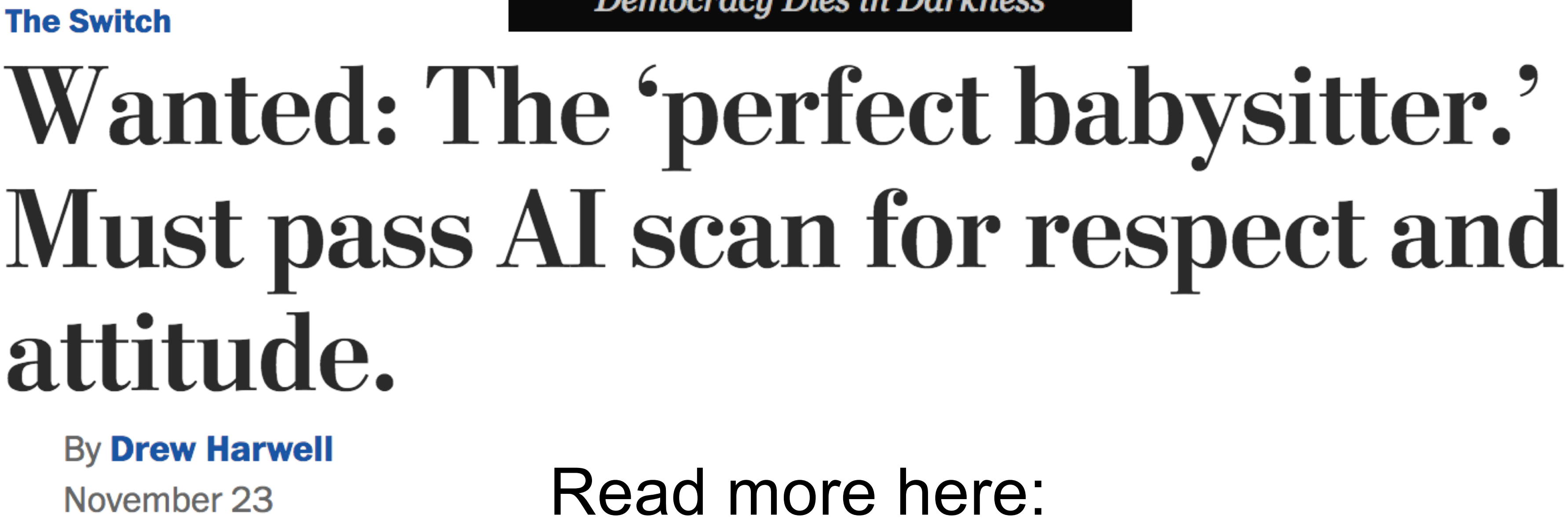
What would we need to make this system sufficiently transparent?

Are concerns with this system solved by explaining outputs?

Possible solutions?

This is not hypothetical.

The Washington Post
Democracy Dies in Darkness



The Switch

Wanted: The ‘perfect babysitter.’ Must pass AI scan for respect and attitude.

By **Drew Harwell**

November 23

Read more here:

https://www.washingtonpost.com/technology/2018/11/16/wanted-perfect-babysitter-must-pass-ai-scan-respect-attitude/

(Break)

Case Two:
Law Enforcement Face Recognition

The police department in Yville wants to be able to identify criminal suspects in crime scene video to know if the suspect is known to detectives or has been arrested before.

Zcorp offers a cloud face recognition API, and the police build a system using this API which queries probe frames from crime scene video against the Yville Police mugshot database.

What does the fact that this is a government application change about the requirements?

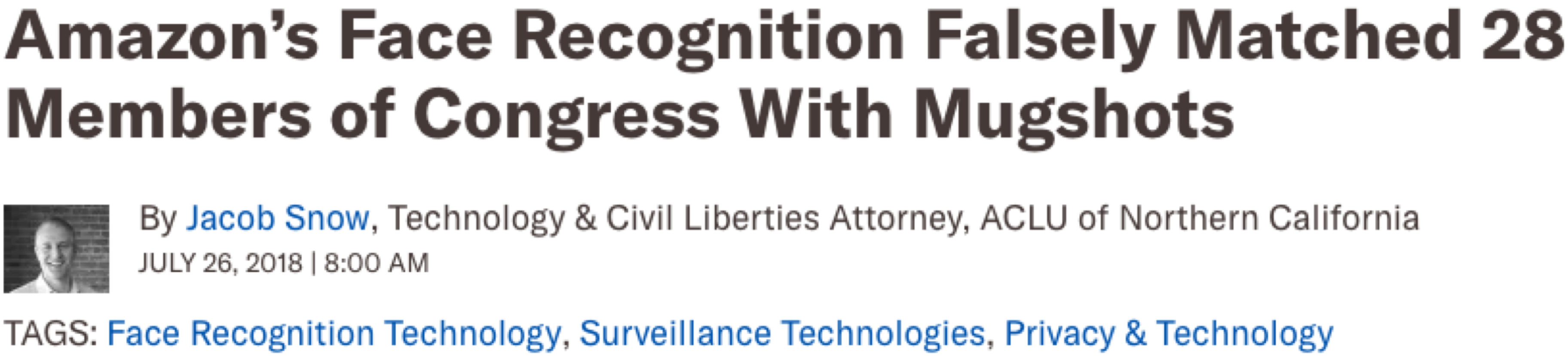
What fairness equities are at stake in such a system?

What is the role of transparency here?

Who has responsibility in or for this system?
What about for errors/mistakes?

What form would explanations take in this system?

This is not hypothetical, either.



Amazon’s Face Recognition Falsely Matched 28 Members of Congress With Mugshots

By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California

JULY 26, 2018 | 8:00 AM

TAGS: Face Recognition Technology, Surveillance Technologies, Privacy & Technology

Read more here:

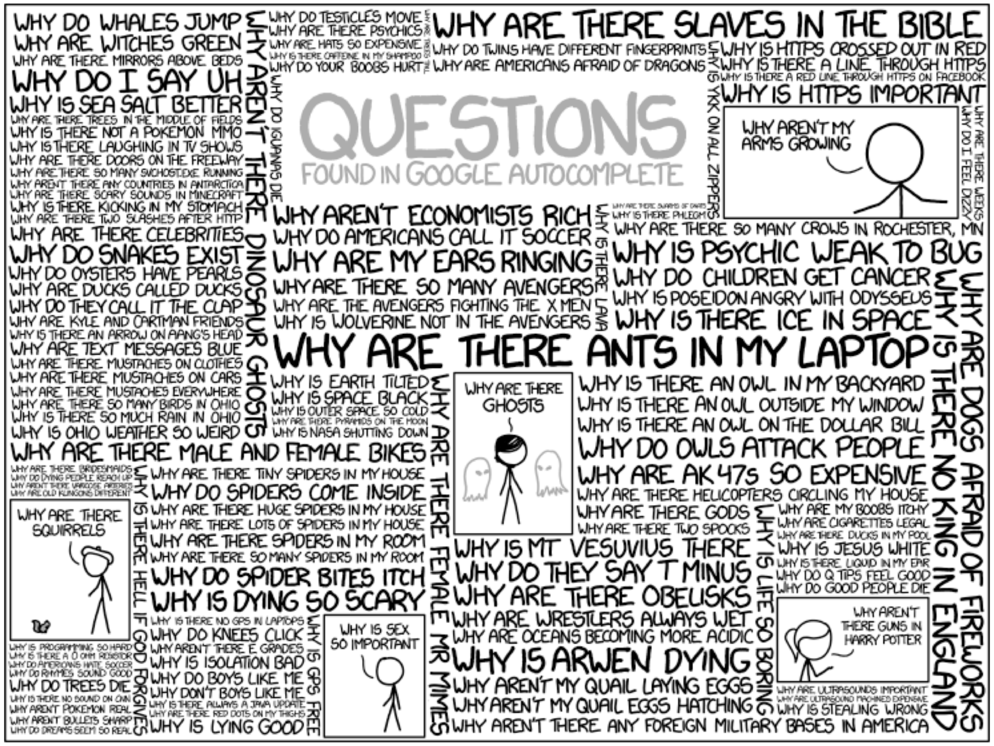
https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

To solve problems with
machine learning, you
must understand them

Respect that others may define the problem differently

If we allow that our systems include people and society,
it’s clear that we have to help ***negotiate*** values,
not simply ***define*** them.

There is no substitute for thinking


QUESTIONS
FOUND IN GOOGLE AUTOCOMPLETE
WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY AREN'T THERE DINOSAUR GHOSTS
WHY DO IGUANAS DIE
WHY DO TESTICLES MOVE
WHY ARE THERE PSYCHICS
WHY ARE HATS SO EXPENSIVE
WHY IS THERE CAFFEINE IN MY SHAMPOO
WHY DO YOUR BOOBS HURT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS YKK ON ALL ZIPPERS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS THERE LAVA
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY ARE MY BOOBS ITCHY
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY IS THERE NO KING IN ENGLAND
WHY ARE DOGS AFRAID OF FIREWORKS
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS THERE HELL IF GOD FORGIVES
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY ARE THERE FEMALE MR MIMES
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS GPS FREE
WHY IS SEX SO IMPORTANT
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY IS LIFE SO BORING
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

Questions?