Protective Optimization Technologies: a proposal for contestation in the world rather than fairness in the algorithm

Seda Gürses
f.s.gurses@tudelft.nl
TPM, TU Delft
COSIC/KU Leuven

AITech Agora

A short introduction...

Name: Seda Gürses

TPM B2.180

Currently: Assistant Professor at MAS Department, TPM

Affiliate at COSIC, Department of Electrical Engineering, KU Leuven

Previously: Post-doc at Princeton University, New York University and KU Leuven

Privacy Engineering

Political Economy of Software Infrastructures and Industries

Optimization Systems and Protective Optimization Technologies

overview

Act I: Going forward, is privacy at stake?

Act II: Optimization systems, a category of its own?

Act III: What can go wrong with optimization?

Act IV: Protective Optimization Technologies?

Act V: Conclusions

Act I

going forward, is privacy what is at stake?

Work in collaboration with Martha Poon, Joris van Hoboken, Femke Snelting

“data is the new oil”?

data compared to a natural resource that can be extracted and exploited

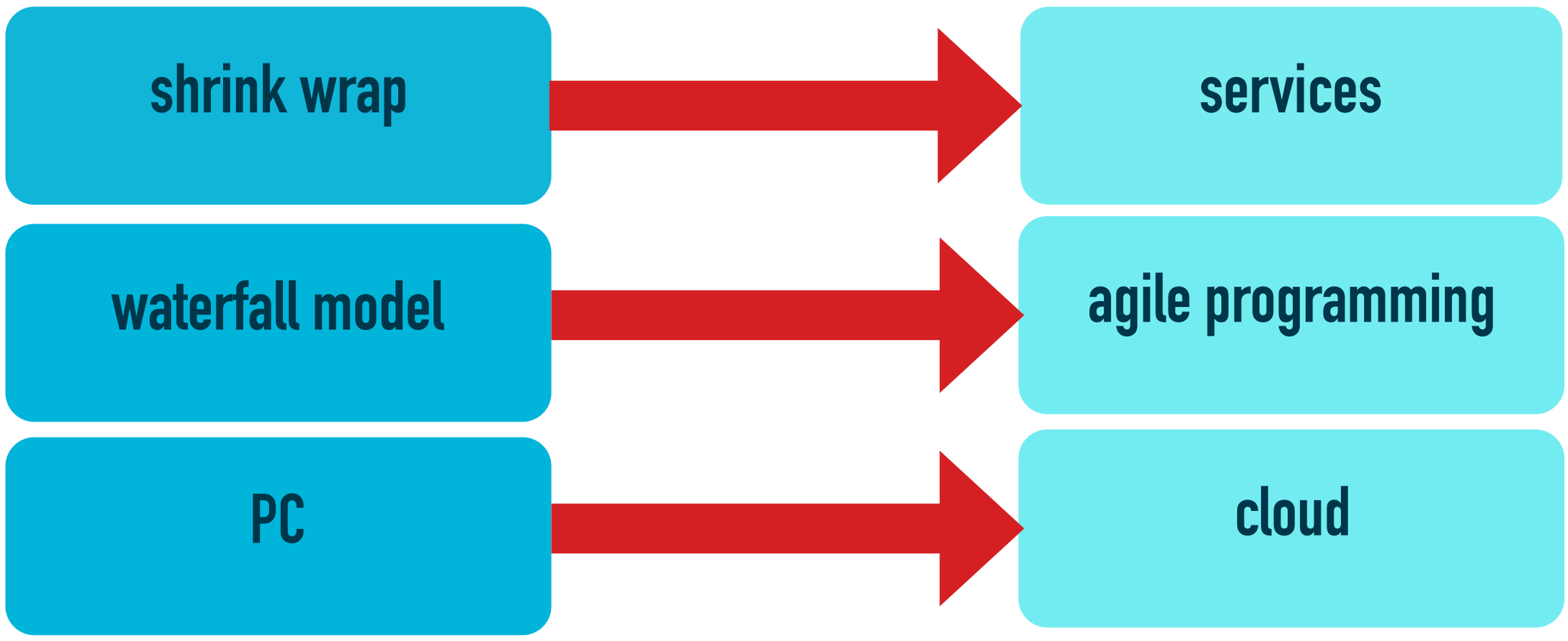
privacy scholars interpret it as “personal data”

data broker industry that guarantees revenue through profiling, targeting ads,

focuses attention on user facing services (consumption) rather than B2B (production) efforts

shrink wrap software

the turn to agile



shrink wrap

enterprise

apps

services

binary runs solely on client side

requires matching soft & hardware

updates & maintenance cumbersome

user has control (oh no!)

pay in advance

Microsoft Word

server (thin) client model

data “secured” by service

updates and maintenance server side

collaborative

pay as you use/trial

office 365

shrink wrap software production

version + purchase

use

time

service bundle

pay per use

use

team integration

SDK/PaaS

cybersecurity

performance

CRM

data brokers

analytics

AB Testing

UX capture

production tools

advertisement

authentication

payment

maps

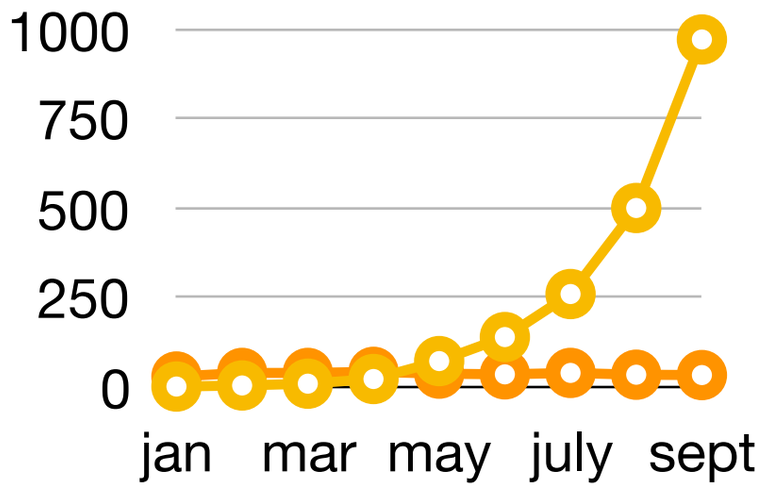
social

embedded media

picture album creation service

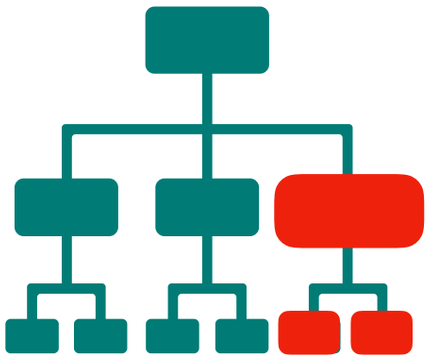
data: more like a lubricant

Computing costs: CapEx -> OpEx



data enables business optimization

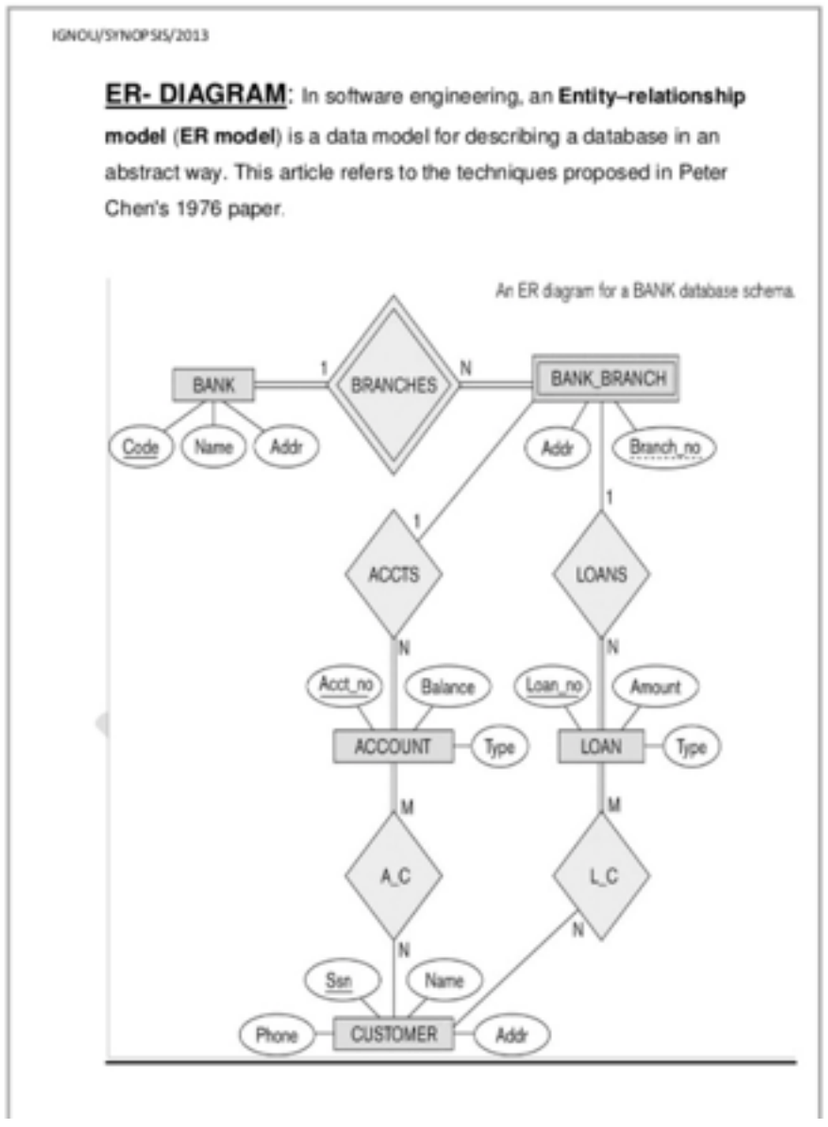
optimization of (computational) resources



agile turn in SE

data enables agile dev

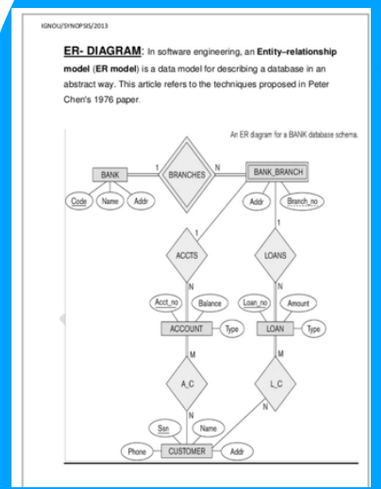
advertisement


IGNOU/SYNOPSIS/2013

ER- DIAGRAM: In software engineering, an **Entity–relationship model** (**ER model**) is a data model for describing a database in an abstract way. This article refers to the techniques proposed in Peter Chen's 1976 paper.

An ER diagram for a BANK database schema.

feedback

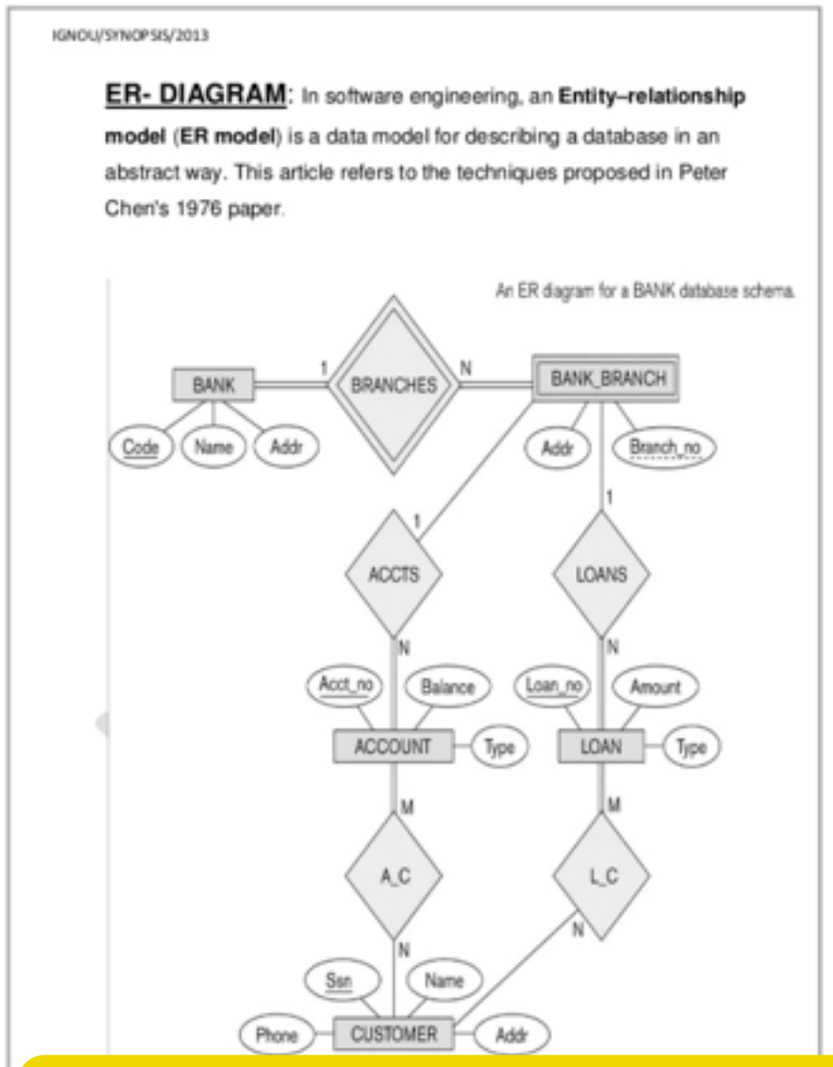

IGNOU/SYNOPSIS/2013

ER- DIAGRAM: In software engineering, an **Entity–relationship model** (**ER model**) is a data model for describing a database in an abstract way. This article refers to the techniques proposed in Peter Chen's 1976 paper.

An ER diagram for a BANK database schema.

- features
- business agility
- business KPIs
- OpEx
- using AI and blockchain

going forward, is privacy what is at stake?


IGNOU/SYNOPSIS/2013

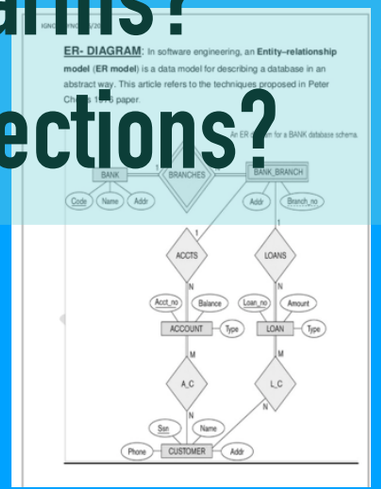
ER- DIAGRAM: In software engineering, an **Entity–relationship model** (**ER model**) is a data model for describing a database in an abstract way. This article refers to the techniques proposed in Peter Chen's 1976 paper.

An ER diagram for a BANK database schema.

information/surveillance/ privacy

optimization harms? protections?

feedback



- features
- business agility
- business KPIs
- OpEx
- using AI and blockchain

Act II

optimization systems, a category of their own?

Work in collaboration with Martha Poon, Joris van Hoboken, Femke Snelting, Carmela Troncoso, Bekah Overdorf, Bogdan

information and communication technologies

optimization systems

optimization systems

capture real- time feedback from users and (operational) environments (cybernetics)

feedback is metricized under the authority of objective functions (optimization)

production and consumption collapsed to enable incremental and adaptive production

capture and manipulate behavior and environments for extraction of value

optimization systems

capture and manipulate behavior and environments for extraction of value

introduce a logic of operational control that focuses on outcomes rather than processes (Poon, 2016)

1. techniques of logistics and control, 2. discourses legitimating a mathematical state as a solution to social contention. (McKelvey, 2018)

collapsing production and consumption often masks labor as a data extraction/computation process

conversion of social, political, cultural, governance issues into economic problems

conflation of allocation of resources with maximization of profit/management of risk.
“consequences of systematic error will be more difficult to observe and control” (Gandy, 2010)

risks and harms

asymmetrical concentration of powers

social sorting

mass manipulation

majority dominance

minority erasure

risks and harms

asymmetrical concentration of powers

optimization systems, a category of their own?

mass manipulation

even if you addressed privacy, these problems could arise!

minority erasure

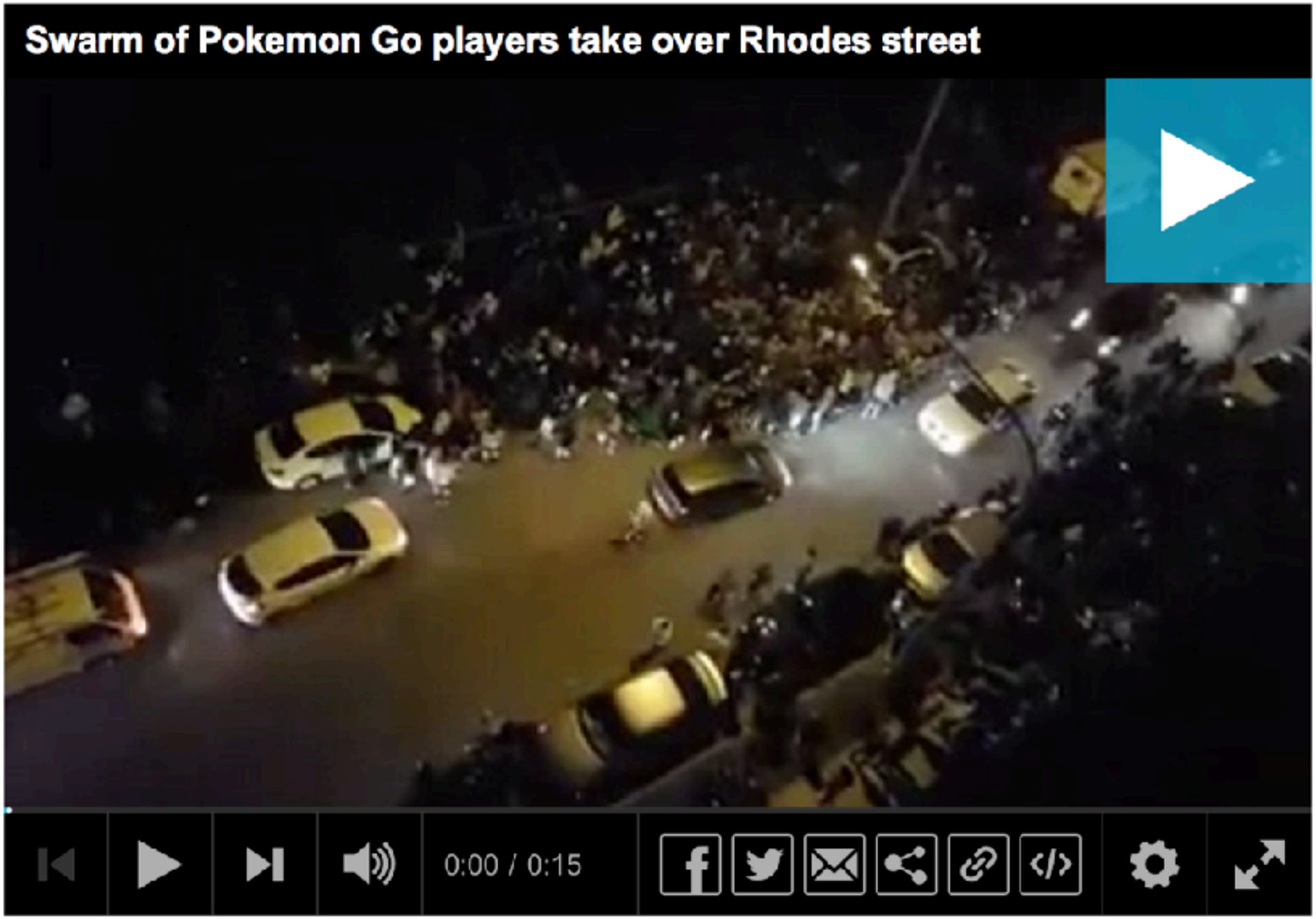
Act III

what could go wrong with optimization?

example: location services

if they are optimizing transport, what is the problem?

co-creation of ideal geographies



http://www.dailymail.co.uk/news/article-3709079/A-road-gridlocked-thousands-Pok-mon-players-swarm-Rhodes-Sydney-street.html

Get the best route, every day, with real–time help from other drivers.

Waze is the world's largest community-based traffic and navigation app. Join other drivers in your area who share real-time traffic and road info, saving everyone time and gas money on their daily commute.

Waze. Outsmarting Traffic, Together.

GET IT ON Google Play Download on the App Store

where that we borrow from traffic engineering and AI safety literature for this

Nothing can beat real people working together

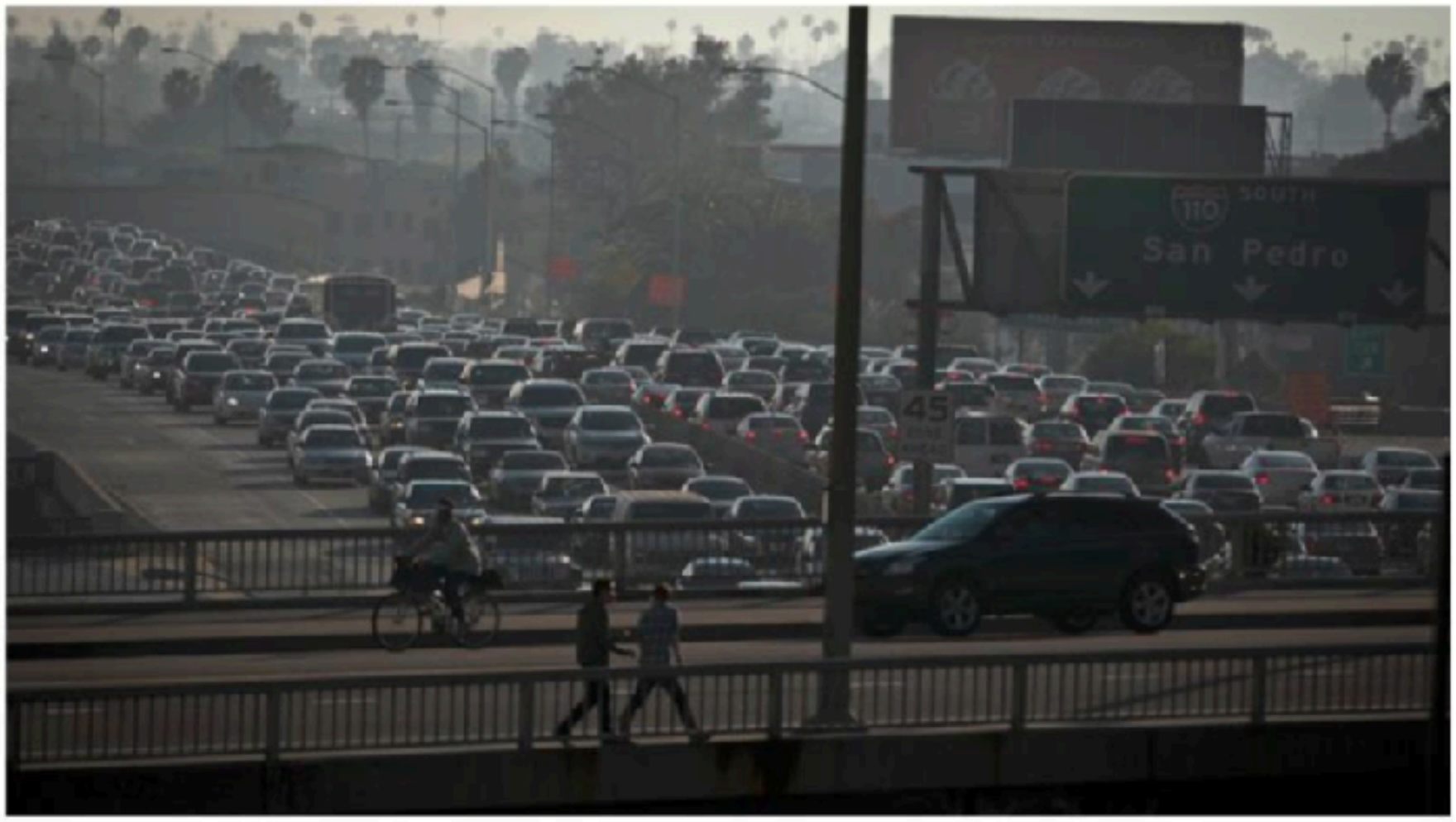
Imagine millions of drivers out on the roads, working together towards a common goal: to outsmart traffic and get everyone the best route to work and back, every day.

TECHNOLOGY

The Perfect Selfishness of Mapping Apps

Apps like Waze, Google Maps, and Apple Maps may make traffic conditions worse in some areas, new research suggests.

ALEXIS C. MADRIGAL MAR 15, 2018



A traffic jam in Los Angeles, like always (REUTERS/BRET HARTMAN)

optimizing for asocial behavior
or negative environmental outcomes

TRAFFIC

Los Angeles councilman tries to work with map apps to alleviate traffic in neighborhoods

EMBED </> MORE VIDEOS ▸

Taking shortcuts around Los Angeles to get to a destination faster is coming at a cost for some fed up locals.

By Veronica Miracle

Wednesday, April 11, 2018

ECHO PARK, LOS ANGELES (KABC) -- Taking shortcuts around Los Angeles to get to a destination faster is coming at a cost for some fed up locals.

disregard non-users

disregard environments

"Without question, the game changer has been the navigation apps... When the primary roads become congested, it directs vehicles into Leonia and pushed them onto secondary roads. We have had days when people can't get out of their driveways."

benefit a few

geles cou
ap apps t
rhoods

Why Some Cities Have Had Enough of Waze

Start-up-turned-tech-giant Waze solves traffic problems for some users, but creates traffic challenges for others.

By **Tala Salem**, Staff Writer May 7, 2018, at 1:42 p.m.

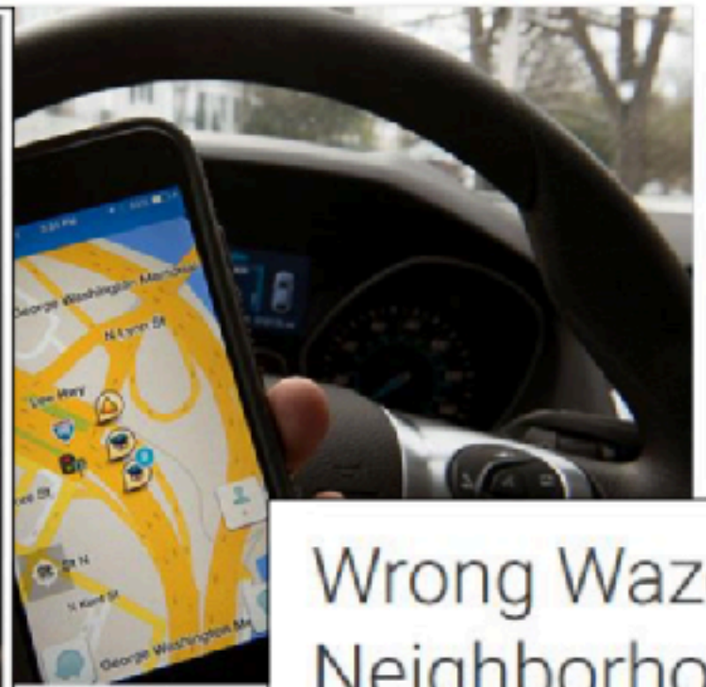


Image Source: http://

ushback for upending local tra
TTY IMAGES

ave caused one New
treme measures

at to expect when he was
ccidents or road constru
hen a different way to wo

g at a cost for some f

Product Reviews + News +

MOBILE

There's a bit of a proble
the Waze navigation a
official claims

SHARE

Community-driven navigation app Waze may be a grea
avoiding traffic jams, but a Los Angeles official claims

Wrong Waze? Residents in San Mateo Irked by Neighborhood Congestion

By NBC Bay Area staff

Published at 7:18 PM PDT on Apr 18, 2018 | Updated at 7:55 PM PDT on Apr 18, 2018

can we identify common externalities of optimization?

disregard non-users and environmental impact

benefit a few

distributional shift

distribution of errors

exploration risks

reward hacking

mass data collection

all while potentially optimizing for asocial behavior
or negative environmental outcomes

can we identify common externalities of optimization?

disregard non-users and environmental impact

benefit a few

fairness

distributional shift

distribution of errors

exploration risks

reward hacking

mass data collection

all while potentially optimizing for asocial behavior
or negative environmental outcomes

problems with fairness frameworks

focuses on a narrow definition of harms in the in the inputs and outputs of an algorithm in a decontextualized manner

proposes to mitigate discrimination harms at the discretion of a service provider that has incentives to optimize otherwise

wherein these service providers exist in an interlocked web of systems that introduce or amplify existing injustices

narrows politics and contestation to the re-design of the algorithm which may not be the site of the problem and may not be the site of the solution

confirms the use of data from (dodgey) data markets without questioning computational power

all of the above cements centralization of (utilitarian) decision making, with some claim to fairness, while reducing opportunities for political contestation

Act IV

Protective Optimization Technologies?

enter POTs

Waze to go: residents fight off crowdsourced traffic… for a while

07 JUN 2016 14

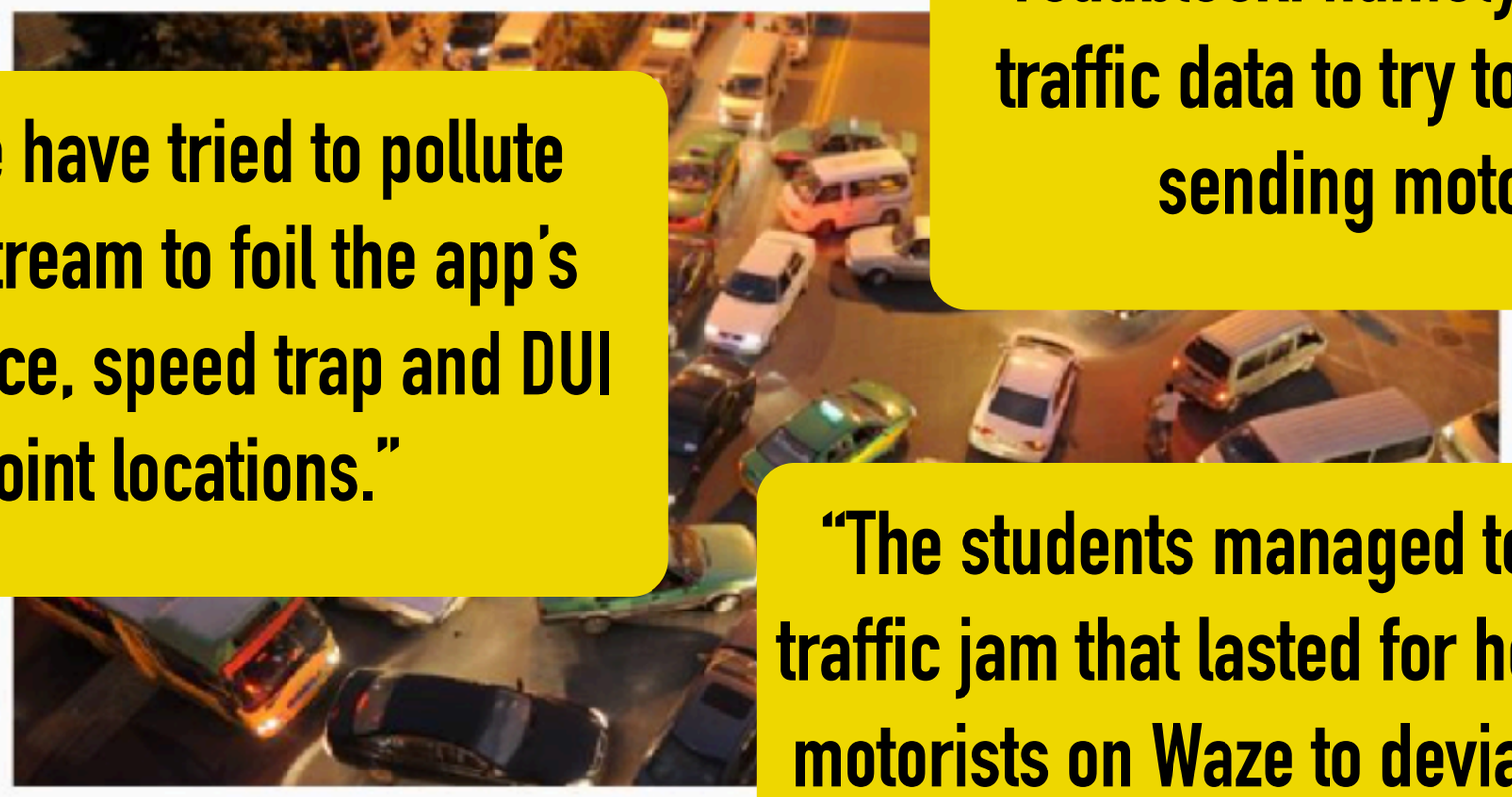
Google, Law & order, Mobile

enter POTs (in the wild)

Waze to go: residents fight off crowdsourced traffic… for a while

07 JUN 2016 14

Google, Law & order, Mobil



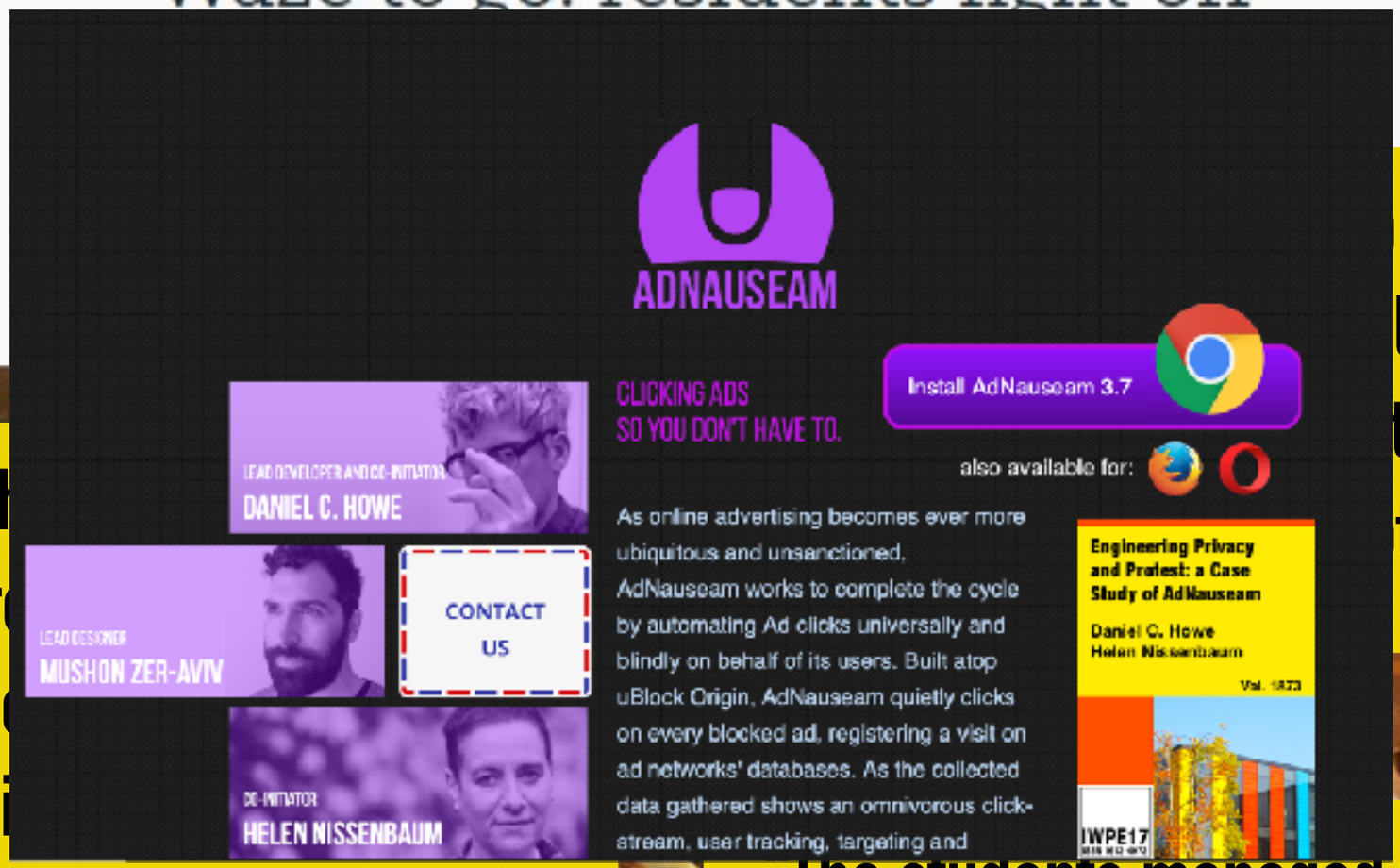
"So he decided to put up his own, virtual roadblock: namely, reporting bogus traffic data to try to trick the app into sending motorists away."

"Miami police have tried to pollute Waze's data stream to foil the app's tracking of police, speed trap and DUI checkpoint locations."

"The students managed to simulate a traffic jam that lasted for hours, causing motorists on Waze to deviate from their planned routes."

enter POTs (in the wild)

Waze to go: residents fight off



"Miami police h
Waze's data str
tracking of polic
checkpoi

ut up his own, virtual
ly, reporting bogus
to trick the app into
torists away."

The students managed to simulate a traffic jam that lasted for hours, causing motorists on Waze to deviate from their planned routes."

Developing POTs

ad-hoc responses: systematize/effectiveness

design tools that allow users to reoptimize themselves and their environment

POTs: when adversarial machine learning meets PETs

optimization systems

capture and manipulate behavior and environments for extraction of value

act I: privacy has become a subproblem

act II: optimization systems are a different beast

act III: optimization systems introduce externalities even if you address (differential) privacy

act IV: we need solutions from the outside (independent of service providers)

Developing POTs: Step 1

Identify externalities

disregard non-users and environmental

benefit a few

distributional

distribution of

exploration risks

reward hacking

mass data collection

all while potentially optimizing for asocial behavior

Developing POTs: Step 2

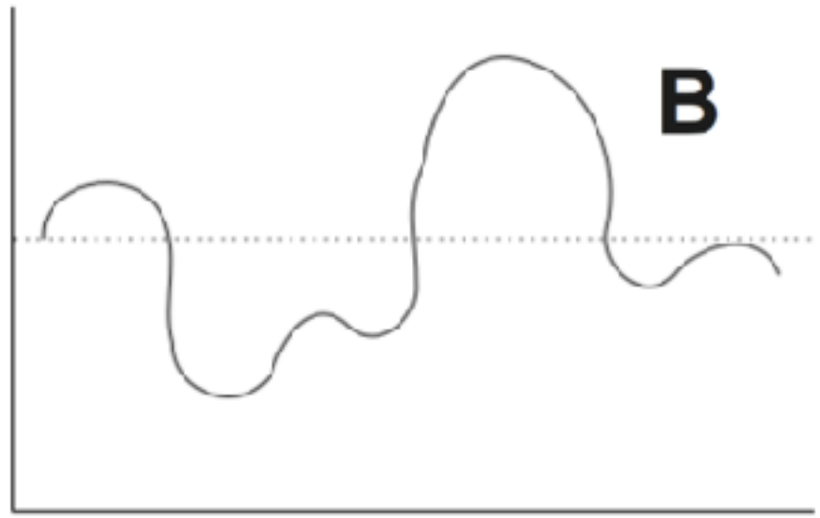
Define a benefit function:

B(X,O)

X: users, non-users, environments

O: observation of system on X

assume low values of B represent externality



Developing POTs

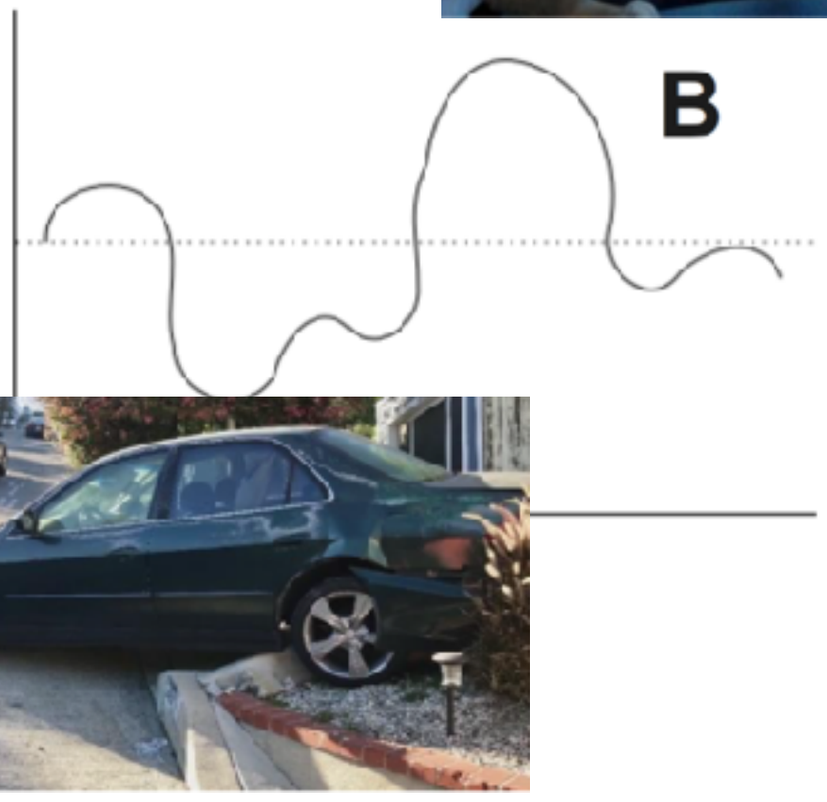
Define a benefit function:

B(X,O)

X: users, non-users, environments

O: observation on X

Look for local minima/negative outcomes!

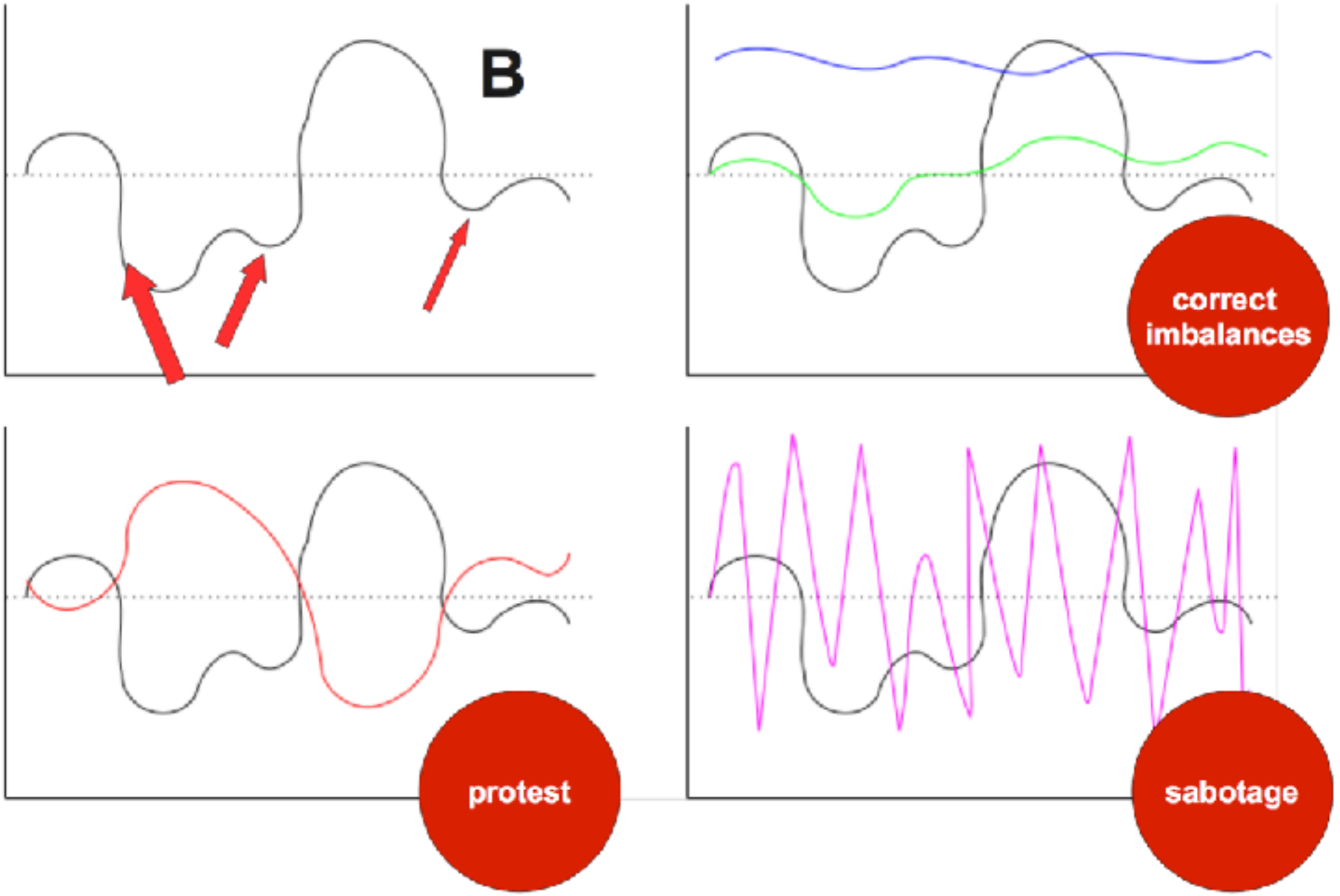


What inputs can you modify?

X -> X'

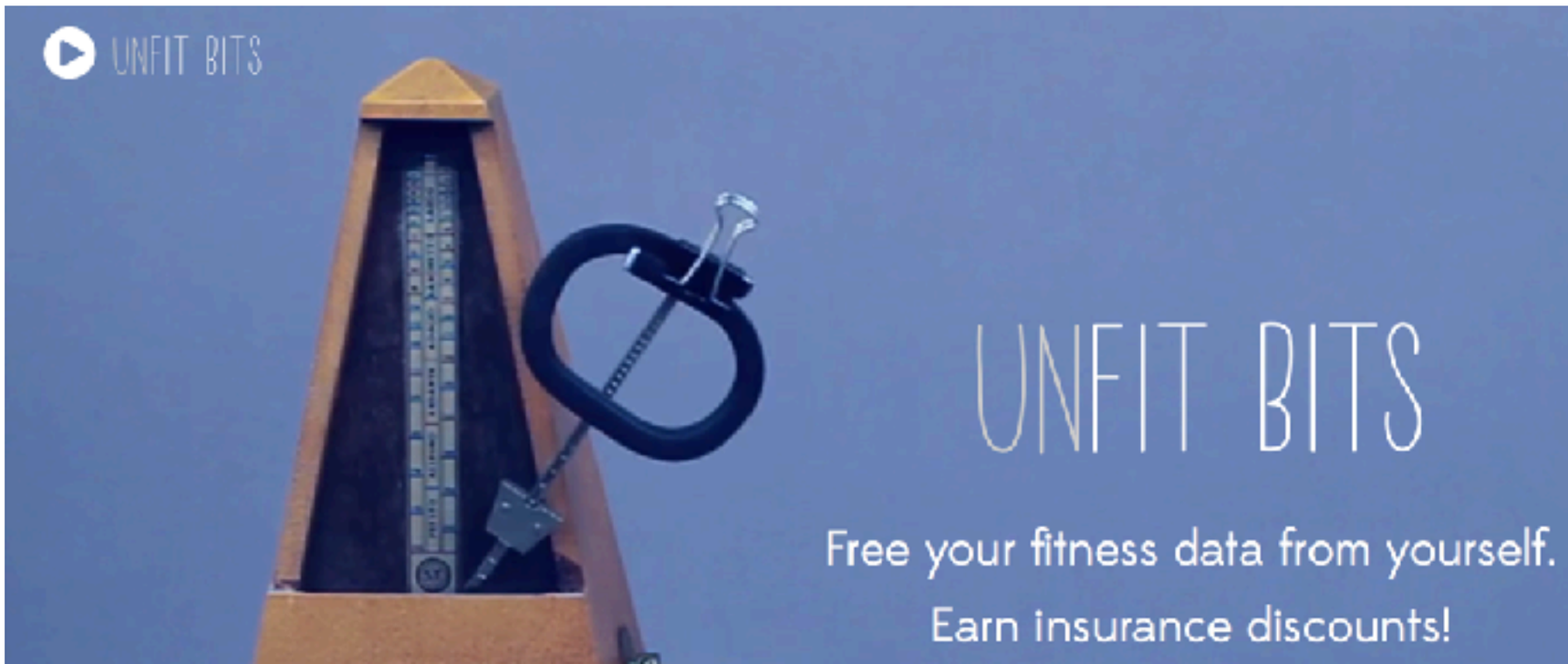
to obtain a desirable O'

Developing POTs



POT in Academia, Arts and the Wild

Origin	Optimization System	Externality	POT	Desired Outcome	Deployme
Academic	Face Recognition	Privacy, discrimination	Wear printed eyeglasses [56]	Evade face detection	Individual
Academic	Copyright Infringement Detection	Fair use takedowns	Adversarial examples [51]	Avoid a fair use takedown	Individual
Academic	Psychometric Profiling	Privacy, manipulations	Text style transfer [57, 58]	Prevent from attribute inference	Individual
Academic	YouTube Recommendations	Manipulation	Poisoning [50]	Breaking out of content bubbles	Individual
Academic	Waze Routing	Local traffic congestion	Sybil devices simulate traffic [62]	Prevent routing into towns	Individual
Academic	GRE Scorer	Biased grading system	Generate essay to pass GRE [63]	Higher test score	Individual
Deployed	Ad Network	Privacy, manipulations	Click on all ads[61]	Ad Network destroyed	Collective
Deployed	Uber Pricing System	Low wages	Shut off app, turn it back on [49]	Induce surge	Collective
Deployed	Instacart Pricing	Low wages	Tip 22¢ in app, cash at door [64]	Fair pay for jobs	Collective
Deployed	Automated Hiring	Bias, discrimination	Edit resume[54]	Flip automated hiring decision	Individual
Deployed	Pokemon Go Resource Spawn	Unfairness	Edit open street maps	Encourage resources to spawn	Individual
Deployed	FitBit for Insurance Premium	Privacy, surveillance	Spoof device location [65]	Get insurance benefits	Individual
Deployed	Pharma Optimizing Patents	End of humanity	Find potential drugs using ML [66]	Get many in the public domain	Individual
Artistic	Face Recognition	Privacy, surveillance	Scarf that is classified as a face [59]	Evade face detection	Individual
Artistic	Face Recognition	Privacy	Camouflage to cover features [53]	Evade face detection	Individual
Artistic	Autonomous Cars	Exploration risks	Ground markings [67]	Trap autonomous cars	Individual



Act V

Conclusions

optimization systems

capture and manipulate behavior and environments for extraction of value

act I: privacy has become a subproblem

act II: optimization systems are a different beast

act III: optimization systems introduce externalities even if you address privacy

act IV: we need solutions from the outside (independent of service providers)

optimization systems

capture and manipulate behavior and environments for extraction of value

what problems are (not) solved with POTs?

POTs as an instance of rethinking trust models and exploring alternative interventions

POTs in service integration (interventions into 3rd party services)

POTs for protection of fundamental rights (Kumar 2018)

when and how are POTs justified? types of pots that are/n't justified?

how can POTs be further formalized?

POTs: are they morally/politically acceptable?

Brunton and Nissenbaum

dishonesty

polluting databases

costs for service providers

costs for other users and environments

more optimization cannot solve optimizations problems

POTs-by-design cannot address all externalities

thank you!

- Philip E. Agre, Surveillance and capture: Two models of privacy, The Information Society, Vol. 10, Iss. 2, 1994 http://steinhardt.nyu.edu/scmsAdmin/uploads/003/648/Agre_SurveillanceAndCapture.pdf
- Oscar Gandy, Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems, Ethics and Information Technology, 2010 https://link.springer.com/article/10.1007/s10676-009-9198-6
- Seda Gürses and Joris Van Hoboken, Privacy After the Agile Turn, Cambridge Handbook of Consumer Privacy, https://www.cambridge.org/core/books/cambridge-handbook-of-consumer-privacy/privacy-after-the-agile-turn/95580B93B4B2446DC5B59166FD2A732F Preprint: https://osf.io/27x3q/
- Irina Kaldrack and Martina Leeker, There is no software, just services, Meson Press, 2015. https://meson.press/wp-content/uploads/2015/06/9783957960566-No-Software-just-Services.pdf
- Martha Poon, Corporate Capitalism and the Growing Power of Big Data: Review Essay, 2016 https://journals.sagepub.com/doi/abs/10.1177/0162243916650491?journalCode=sthd
- Bogdan Kulynych et al. Protective Optimization Technologies, https://arxiv.org/pdf/1806.02711.pdf 2019
- Rebekah Overdorf et al., Questioning the assumptions behind fairness solutions, CoRR, 2018, https://arxiv.org/abs/1811.11293