Learning from the People

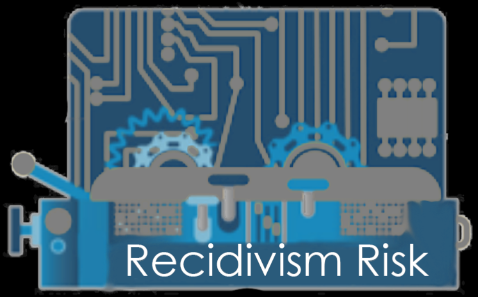
From Normative to Descriptive Solutions to Problems in Security, Privacy & Machine Learning

Elissa M. Redmiles, Microsoft Research & Max Planck Institute for Software Systems

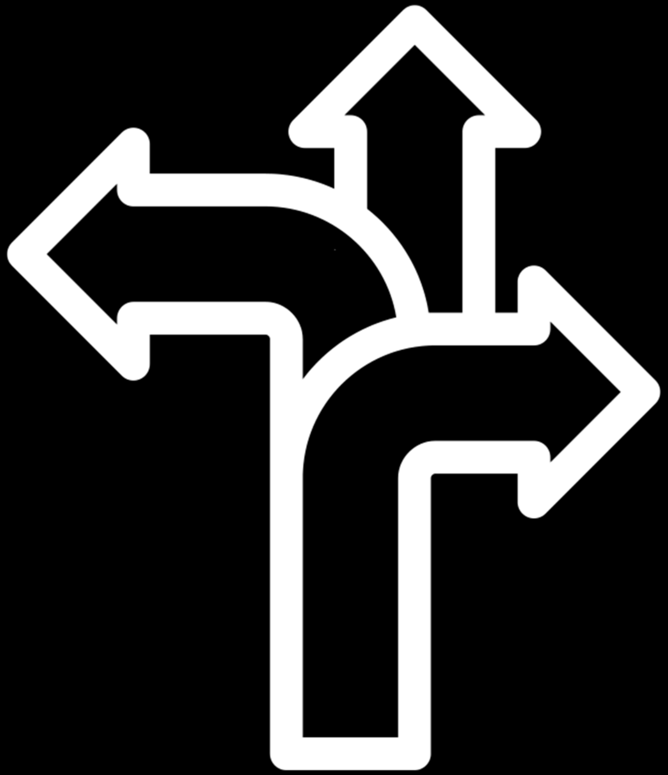
@eredmil1

eredmiles@gmail.com

Computational problems require constant decision-making

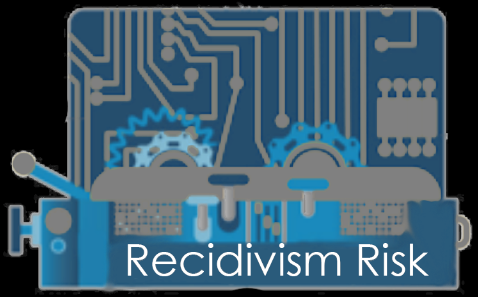


Which Features
Are Fair to Use?



Which Security
Requirements to Set?

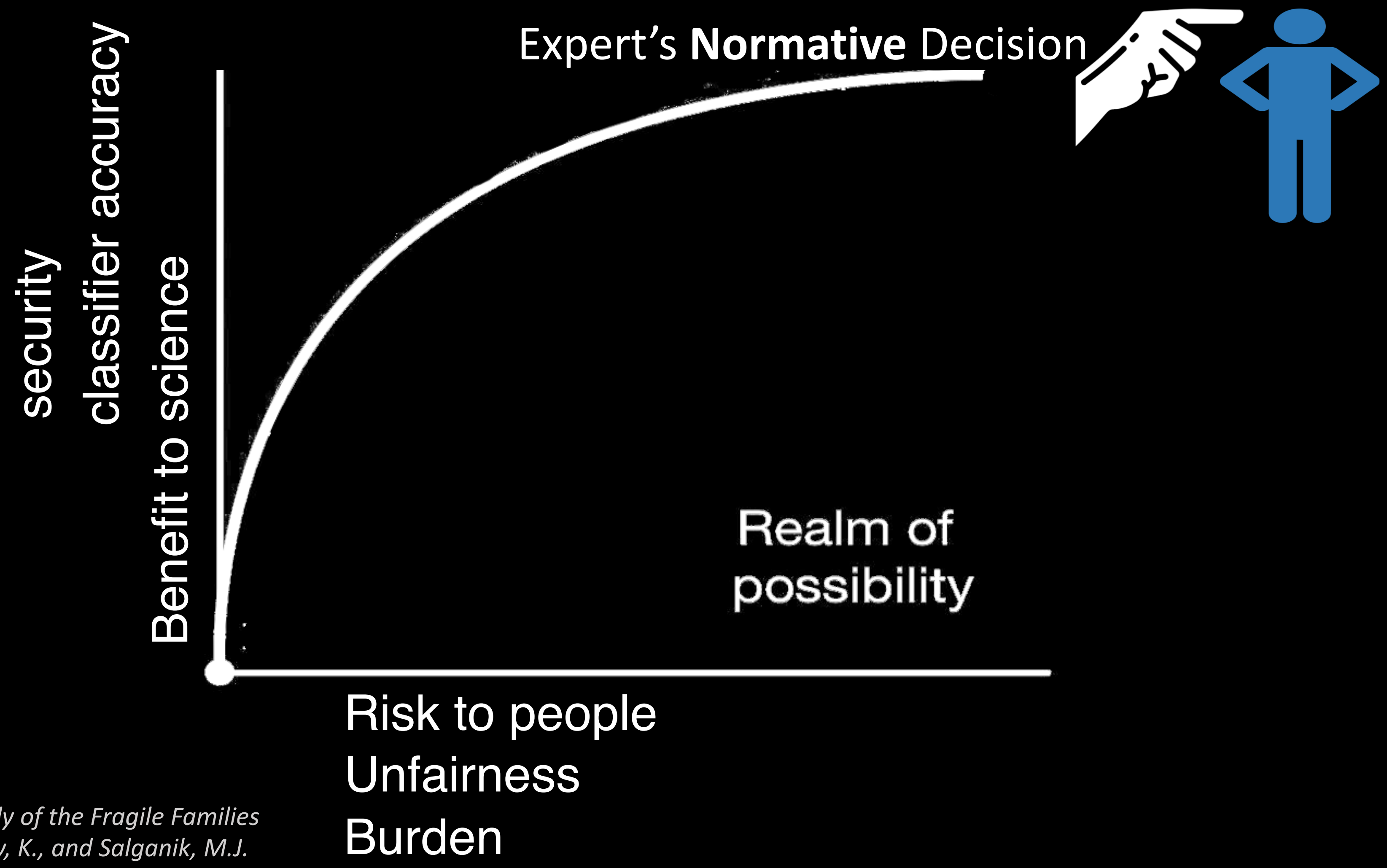
Typically: experts set best practices



Which Features
Are Fair to Use?

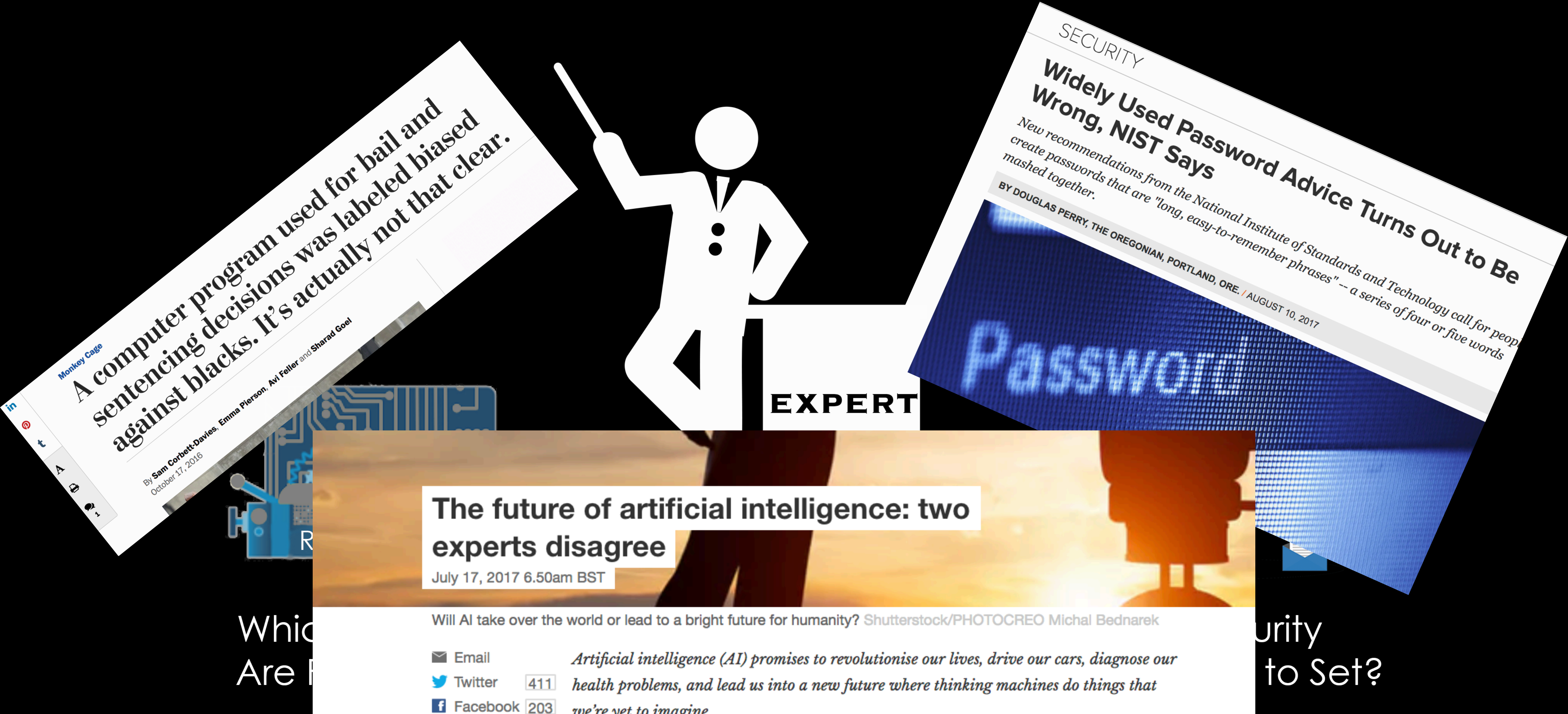
Which Security
Requirements to Set?

Experts trade off costs and benefits



Original (white) figure credit:
Privacy, ethics, and data access: A case study of the Fragile Families Challenge. Lundberg, I., Narayanan, A., Levy, K., and Salganik, M.J.

Experts do not always agree on best practices



Monkey Cage

A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear.

By Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel

October 17, 2016

SECURITY

Widely Used Password Advice Turns Out to Be Wrong, NIST Says

New recommendations from the National Institute of Standards and Technology call for peopl create passwords that are "long, easy-to-remember phrases" -- a series of four or five words mashed together.

BY DOUGLAS PERRY, THE OREGONIAN, PORTLAND, ORE. / AUGUST 10, 2017

The future of artificial intelligence: two experts disagree

July 17, 2017 6.50am BST

Will AI take over the world or lead to a bright future for humanity? Shutterstock/PHOTOCREO Michal Bednarek

Email

Twitter 411

Facebook 203

Artificial intelligence (AI) promises to revolutionise our lives, drive our cars, diagnose our health problems, and lead us into a new future where thinking machines do things that we're yet to imagine.

R

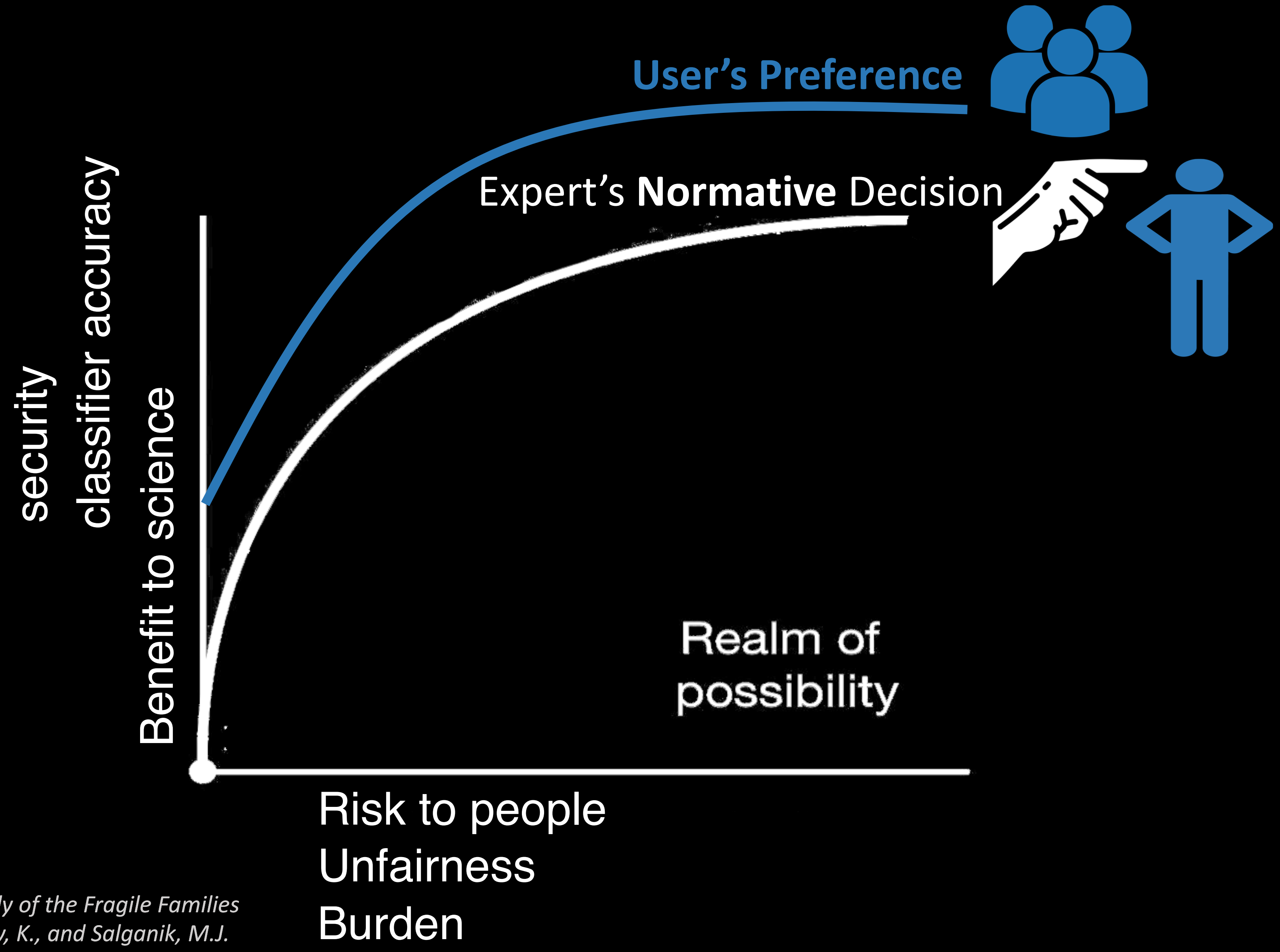
Whic

Are F

urity

to Set?

More importantly, users and experts may disagree

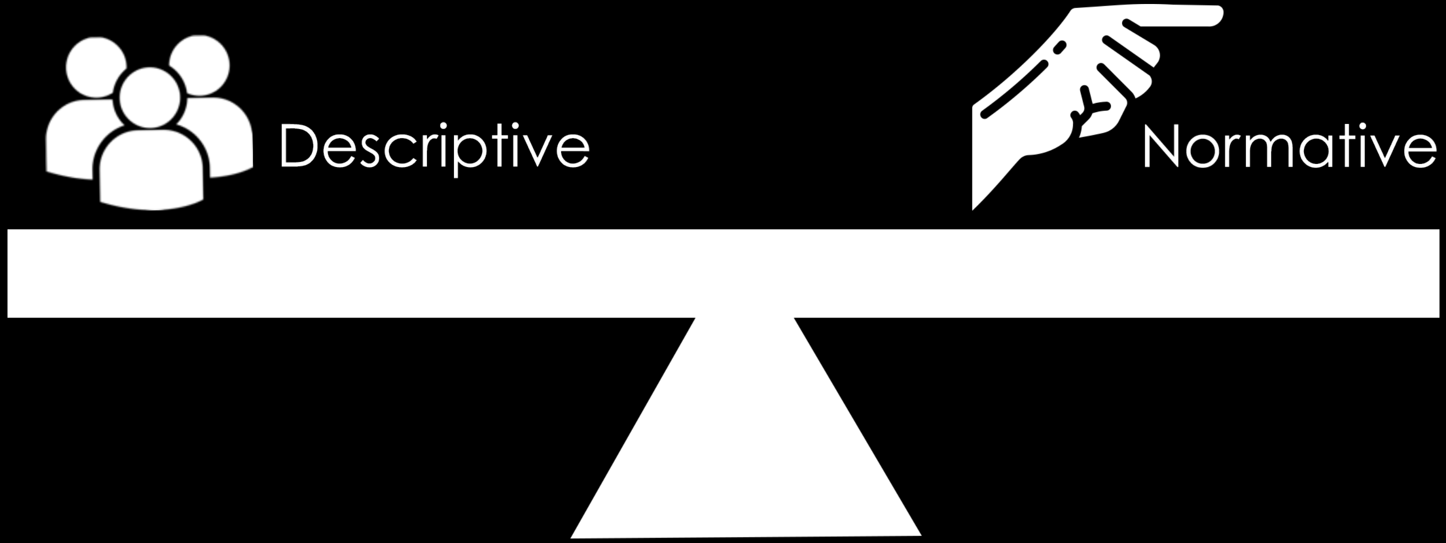


Original (white) figure credit:
Privacy, ethics, and data access: A case study of the Fragile Families Challenge. Lundberg, I., Narayanan, A., Levy, K., and Salganik, M.J.

This disagreement is a classic tension in moral philosophy

Normative	Descriptive
Experts prescribe best practices	**Learn** non-expert preference/behavior **Infer** best practices

Can we use descriptive approaches in computational decision-making?



Descriptive

Normative

Three case studies, three different descriptive methods

How should we set **security** policies?

- **Observe** behavior
- Infer preference
- Make decision

Which features are fair to use in **machine learning**?

- **Ask** preference
- Make decision

What content should be allowed in virtual reality?

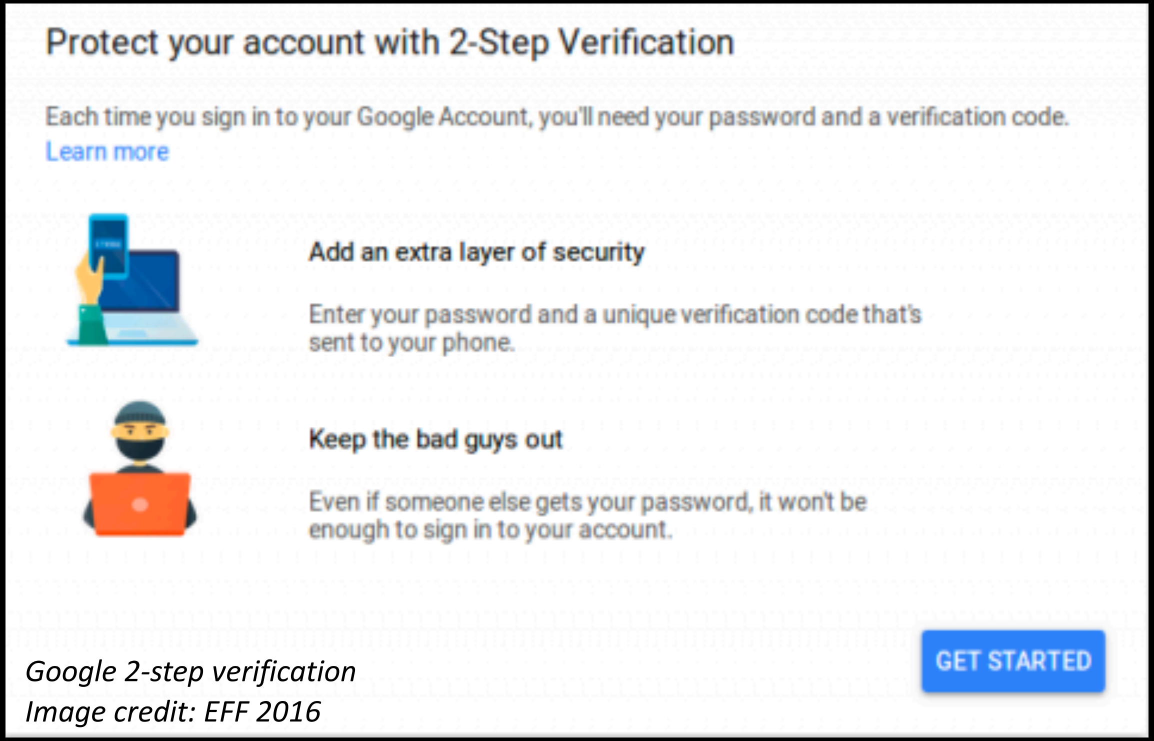
- Make decision **together**

Security

Determine how & when to prompt secure behavior

Goal

Get users to behave more securely by prompting



Protect your account with 2-Step Verification

Each time you sign in to your Google Account, you'll need your password and a verification code. Learn more

Add an extra layer of security

Enter your password and a unique verification code that's sent to your phone.

Keep the bad guys out

Even if someone else gets your password, it won't be enough to sign in to your account.

Google 2-step verification
Image credit: EFF 2016

Why don't users behave as expected when prompted?

The user is going to pick
dancing pigs over **security** every time.

-- McGraw and Felten / Schneier

Measure prompt response using a novel, scalable behavioral-economics security experimentation system

Online experimental system: simple bank account

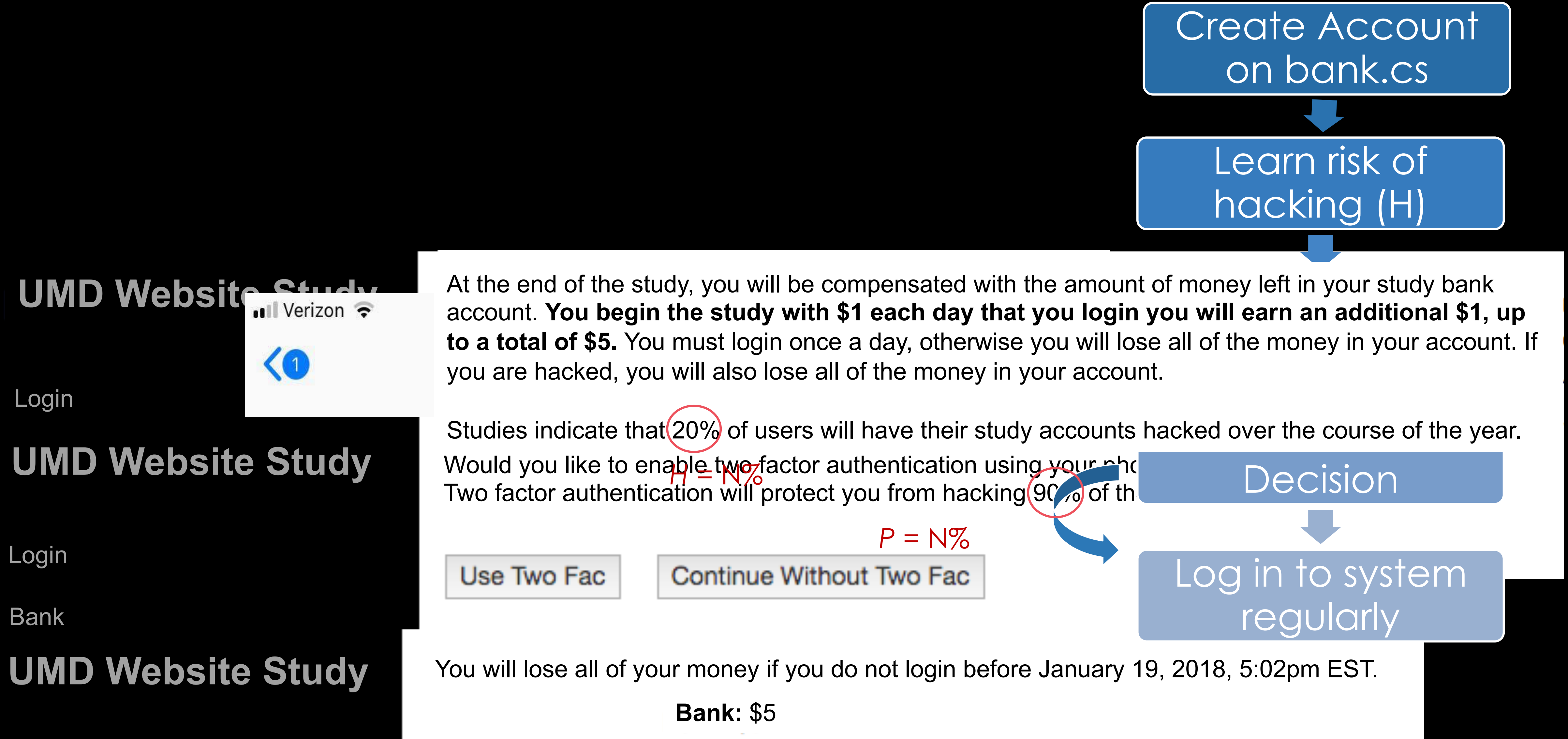
Account holds study compensation

- Account has explicit **risk** of being hacked

Redmiles, E.M., Mazurek, M.L., and Dickerson, J.P. *Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions.* ACM **Economics & Computation (EC)** 2018.

Featured on
Schneier on Security
Blog Newsletter Books Essays News Talks Academic

Participants interact with simulation system
We observe their responses to security prompts



Create Account on bank.cs

Learn risk of hacking (H)

Decision

Log in to system regularly

At the end of the study, you will be compensated with the amount of money left in your study bank account. **You begin the study with $1 each day that you login you will earn an additional $1, up to a total of $5.** You must login once a day, otherwise you will lose all of the money in your account. If you are hacked, you will also lose all of the money in your account.

Studies indicate that 20% of users will have their study accounts hacked over the course of the year.

Would you like to enable two factor authentication using your pho

Two factor authentication will protect you from hacking 90% of th

Use Two Fac

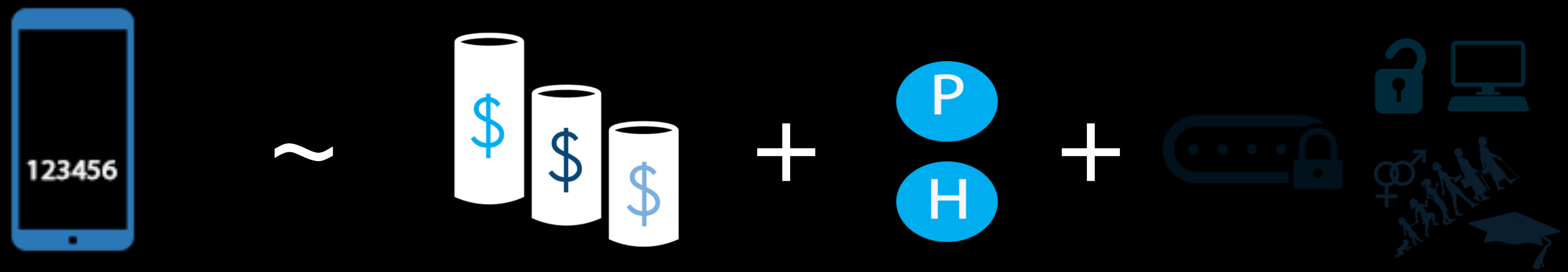
Continue Without Two Fac

You will lose all of your money if you do not login before January 19, 2018, 5:02pm EST.

Bank: $5

Only 52% of participants enabled 2FA.

Testing the bounded rationality hypothesis: is there a consistent pattern in security behavior?



Enable 2FA

Account value

Risk
with/out 2FA

Controls
Password Strength
Internet & Security Skill
Demographics
(Gender, Age, Education)

Neural Net Strength Meter
Ur et al. 2017

Validated Scales
Hargittai & Hsieh 2013
Egelman & Peer 2015

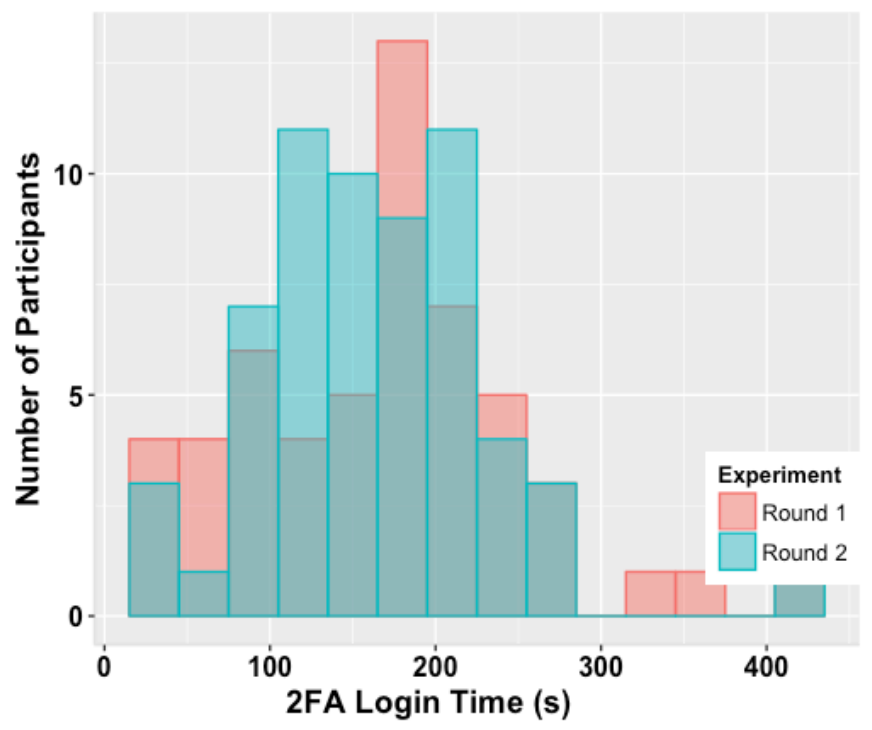
Redmiles, E.M., Mazurek, M.L., and Dickerson, J.P. *Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions.* Economics & Computation (**EC2018**).

Testing the bounded rationality hypothesis: is there a consistent pattern in security behavior?

123456

Enable 2FA

~



Costs

proxy:
time spent

+

123456

Past Behavior

(RD1 2FA choice)

+

Controls

Redmiles, E.M., Mazurek, M.L., and Dickerson, J.P. *Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions.* Economics & Computation (**EC2018**).

Experimental results suggest users are boundedly rational

H

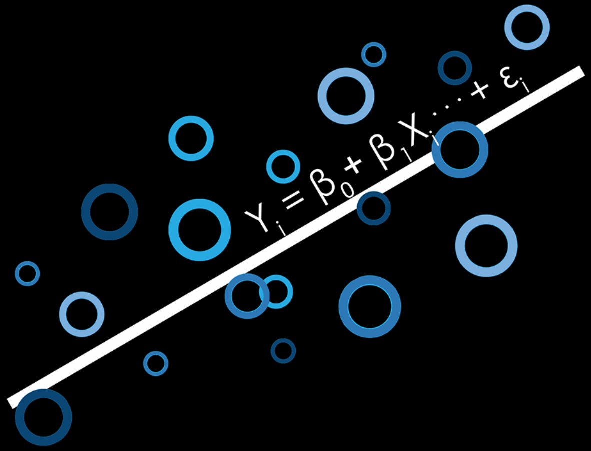
P

Risk (H, P) + **Account Value** (Earn/Endow)

explains 9% behavior variance

Redmiles, E.M., Mazurek, M.L., and Dickerson, J.P. *Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions.* Economics & Computation (**EC2018**).

Behavior is explainable
Differences in ability and account value alter behavior



People behave in ways we can model well

We can model human behavior well (R^2=0.61) as a function of variables measured or controlled in the simulation system

Differences in *ability* (differences in *cost*) alter behavior

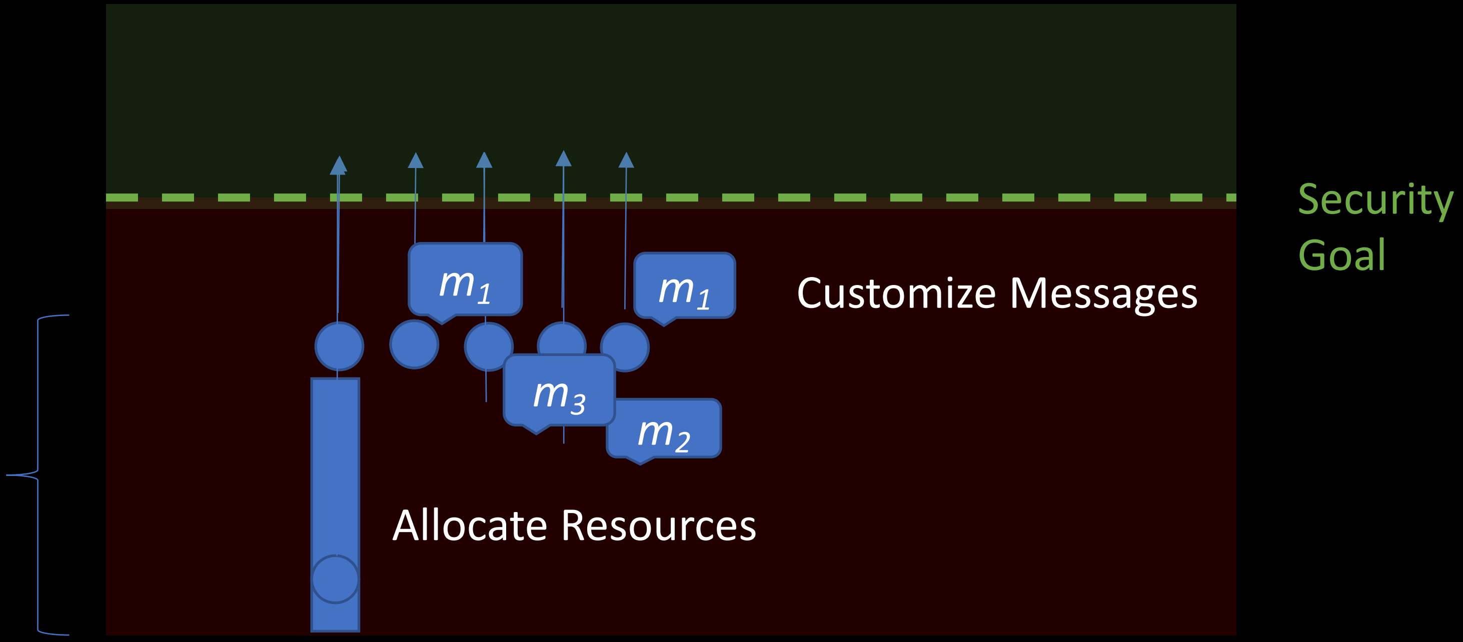
Differences in *account valuation* alter behavior

Redmiles, E.M., Mazurek, M.L., and Dickerson, J.P. *Dancing Pigs or Externalities? Measuring the Rationality of Security Decisions.* Economics & Computation (**EC2018**).

Normative

Prompt everyone to use 2FA until they do: it's good for them

- Problem: people are so inundated they start ignoring prompts
- Problem: not everyone gets the same value out of the same behavior



Inequalities in Ability (e.g., 2FA difficulty)
Valuation of account
…

Can we use our descriptive knowledge to set prompts?

How should we set **security** policies?

- **Observe** behavior ✓
- Infer preference ✓
- Make decision

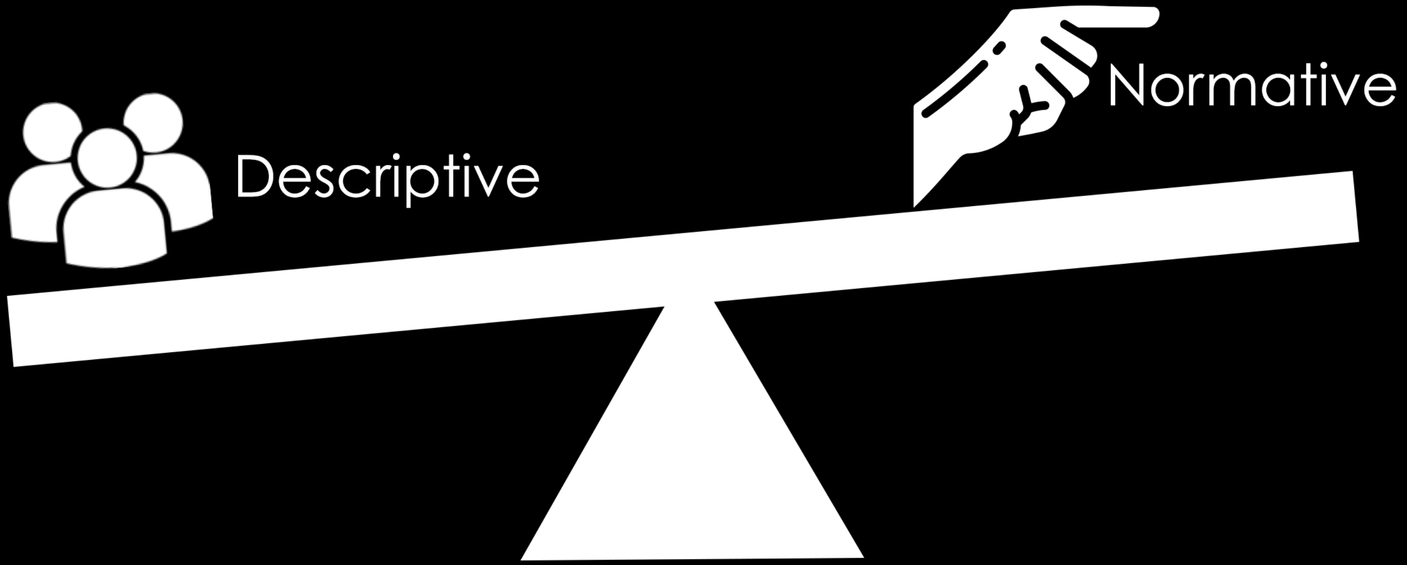
Which features are fair to use in **machine learning**?

- **Ask** preference
- Make decision

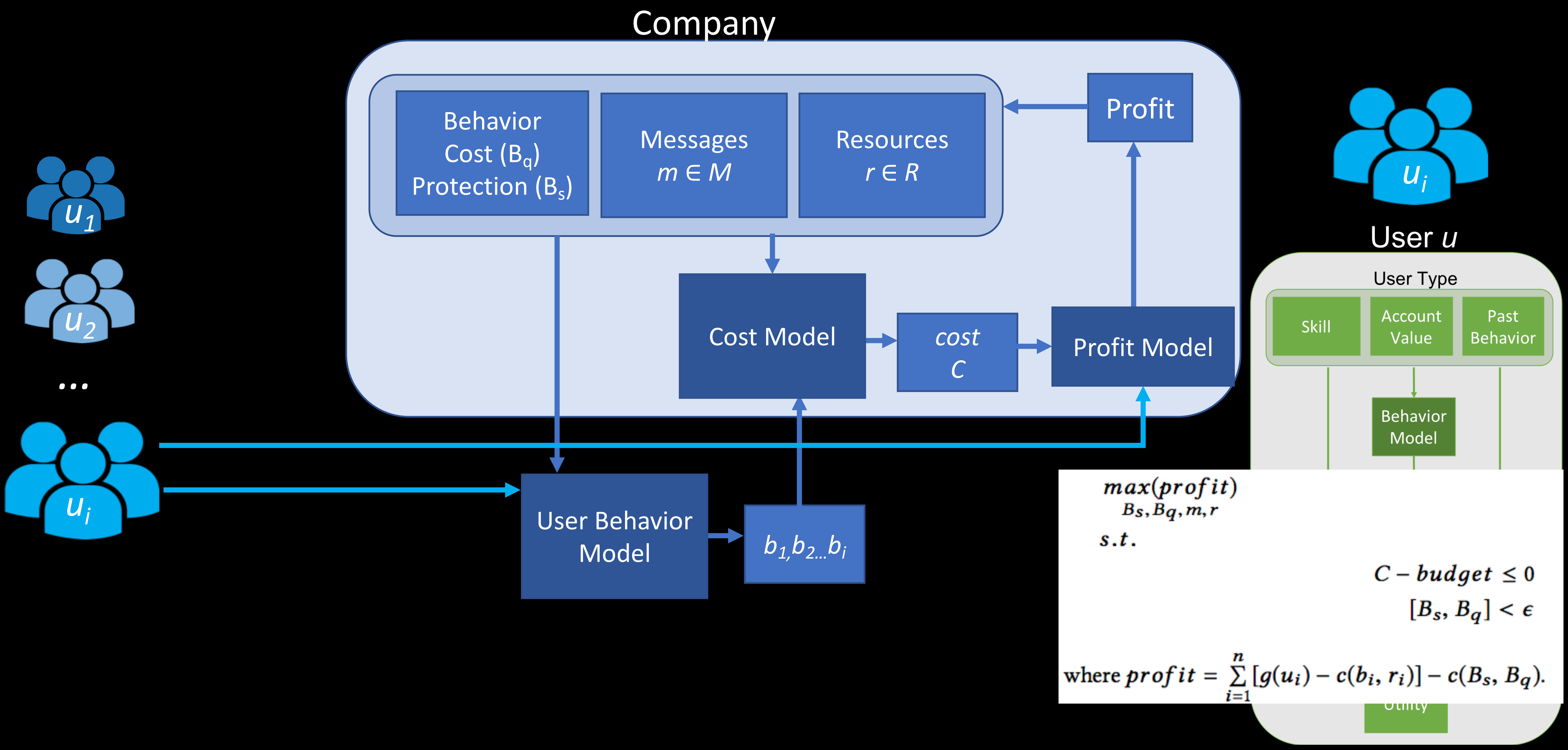
What content should be allowed in virtual reality?

- Make decision **together**

Mechanism design to facilitate descriptive approach



Companies can maximize profit by selecting optimal values for factors they control



$$\max_{B_s, B_q, m, r}(profit)$$

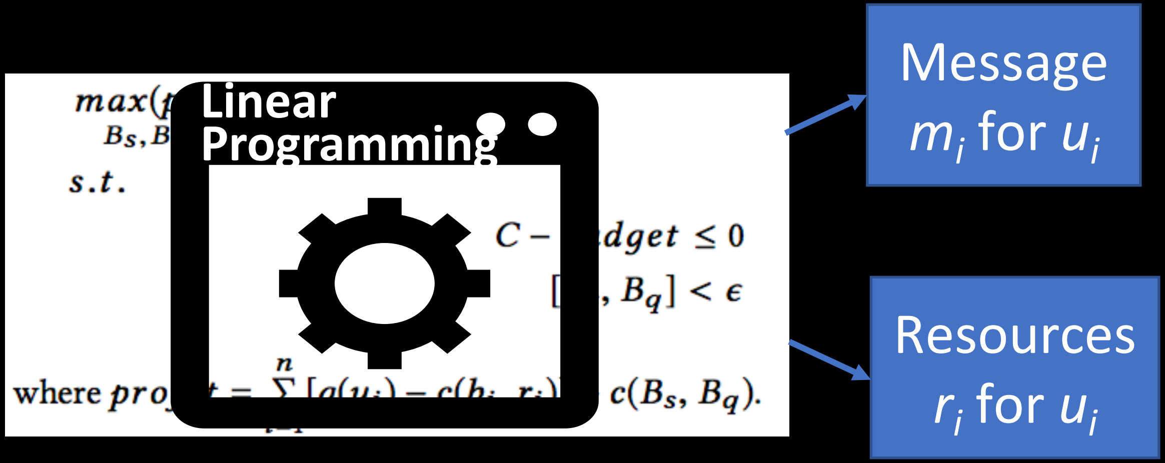
s.t.

$$C - budget \leq 0$$

$$[B_s, B_q] < \epsilon$$

where $profit = \sum_{i=1}^{n} [g(u_i) - c(b_i, r_i)] - c(B_s, B_q)$.

Mechanism design enables descriptive approach and introduction of equity notions



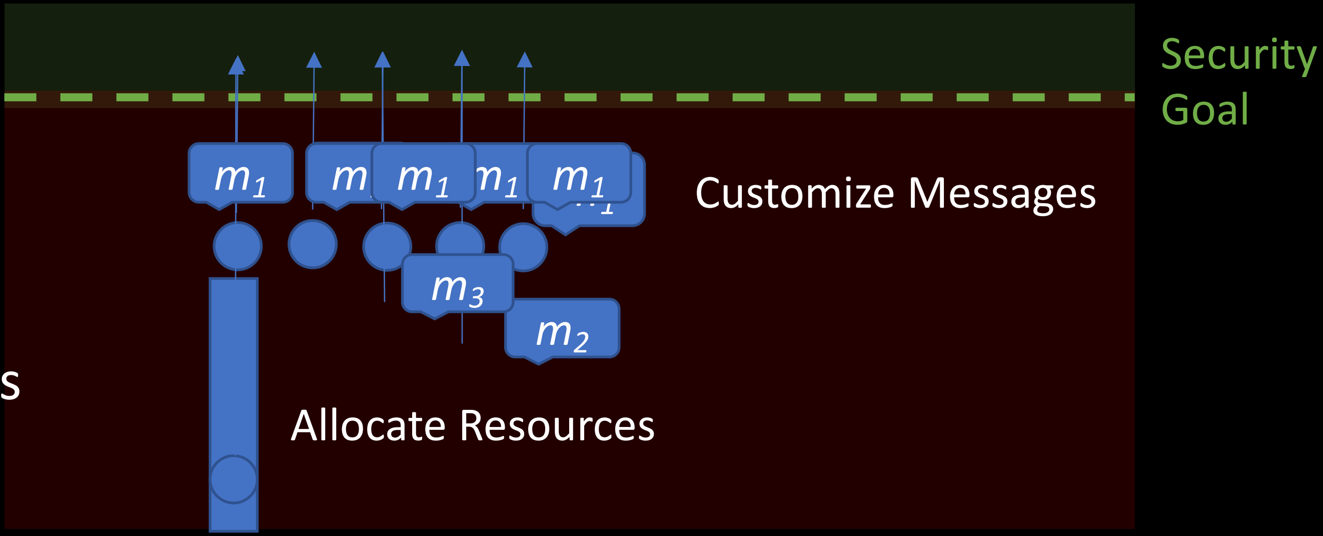
Constraints

Inequalities in Ability (e.g., 2FA difficulty)

Valuation of account

…

- Effort equity: minimize variance in costs
- Risk equity: minimize variance in risk



Decide by solving an optimization problem that uses knowledge of user behavior gained through observation

How should we set **security** policies?

- **Observe** behavior ✓
- Infer preference ✓
- Make decision ✓

Which features are fair to use in **machine learning**?

- **Ask** preference
- Make decision

What content should be allowed in virtual reality?

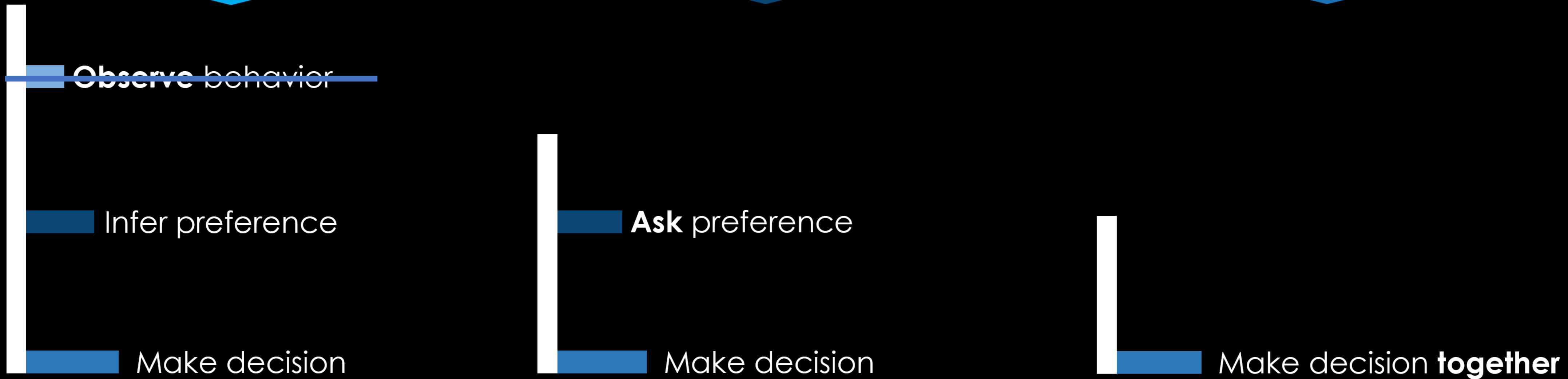
- Make decision **together**

Can we get to a decision sooner? Directly ask the users

How should we set **security** policies?

Which features are fair to use in **machine learning**?

What content should be allowed in virtual reality?



~~**Observe** behavior~~

Infer preference

Make decision

Ask preference

Make decision

Make decision **together**

Machine Learning

Select features that are fair to use for classification

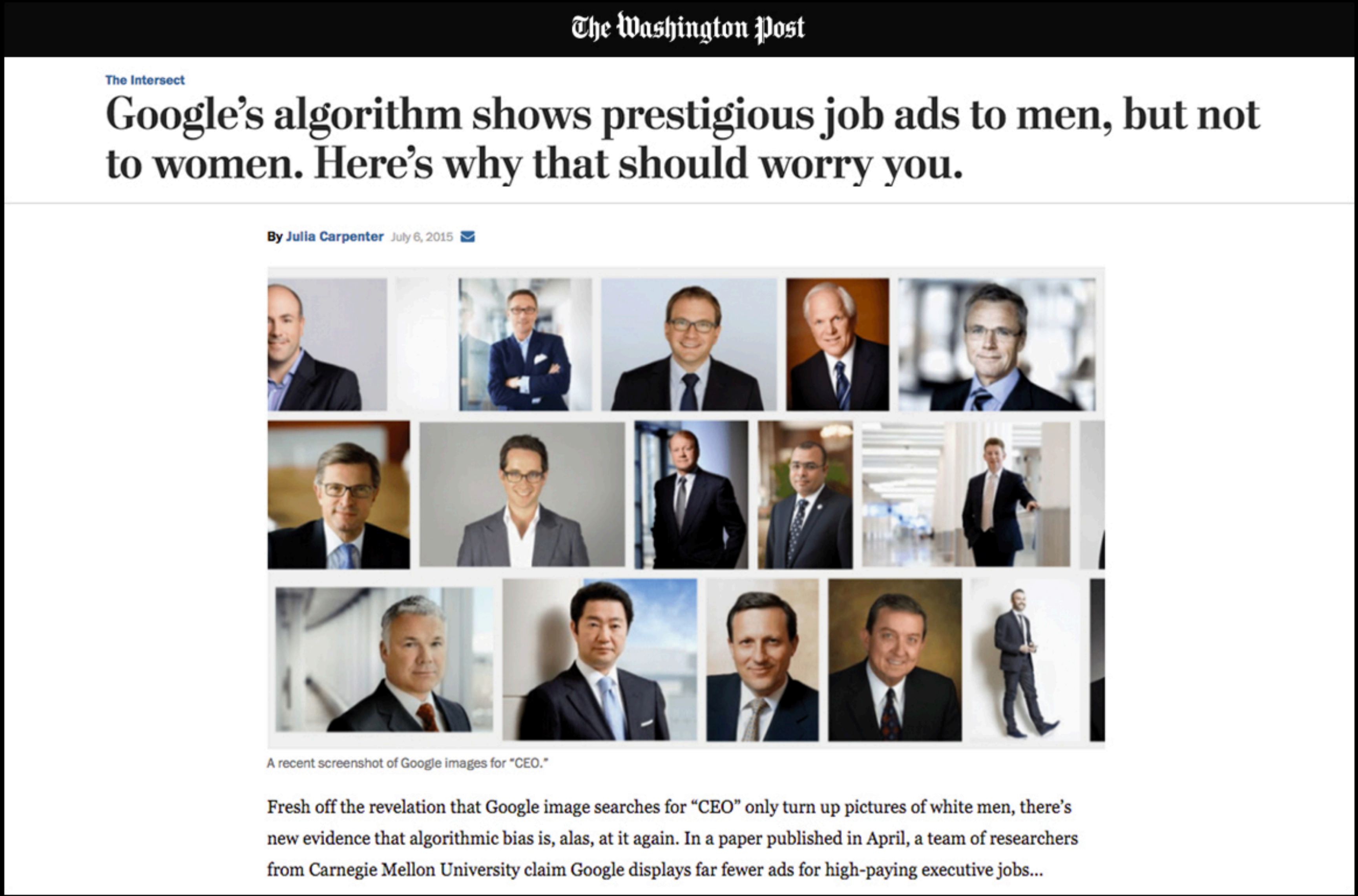
Goal

Determine which features are fair to use in a classifier

Descriptive Approach

Model how users reason about fairness and include/weight features based on fairness judgements

Let’s back up for a moment:
why do we care about feature fairness?



The Washington Post

The Intersect

Google’s algorithm shows prestigious job ads to men, but not to women. Here’s why that should worry you.

By Julia Carpenter July 6, 2015

A recent screenshot of Google images for “CEO.”

Fresh off the revelation that Google image searches for “CEO” only turn up pictures of white men, there’s new evidence that algorithmic bias is, alas, at it again. In a paper published in April, a team of researchers from Carnegie Mellon University claim Google displays far fewer ads for high-paying executive jobs...

What drives perceptions of ad discrimination scenario?

Systemy is a local technology firm that develops software. They are expanding and want to hire new employees. Systemy contracts with Bezo Media, an online advertising network, which places Systemy's job ad on a local news website. Bezo Media charges less to reach individuals who are Asian than who are White. Bezo Media's marketing computer program automatically selects the less expensive option.

Systemy requests that the ad be shown to viewers who have recently visited technology-interest websites. Individuals who are Asian tend to visit more technology-interest websites than individuals who are White.

[illegible] rather than individuals who are White.

As a result, the ad is shown more frequently to individuals who are Asian than who are White.

Plane, A., Redmiles, E.M., Mazurek, M.L., and Tschantz, M. *Exploring User Perceptions of Discrimination in Online Targeted Advertising.* **USENIX Security 2017**.

Measured the effect of varying beneficiary, targeting mechanism & targeted features

Training Data Collection

MTurk survey (n=191) for training regression models

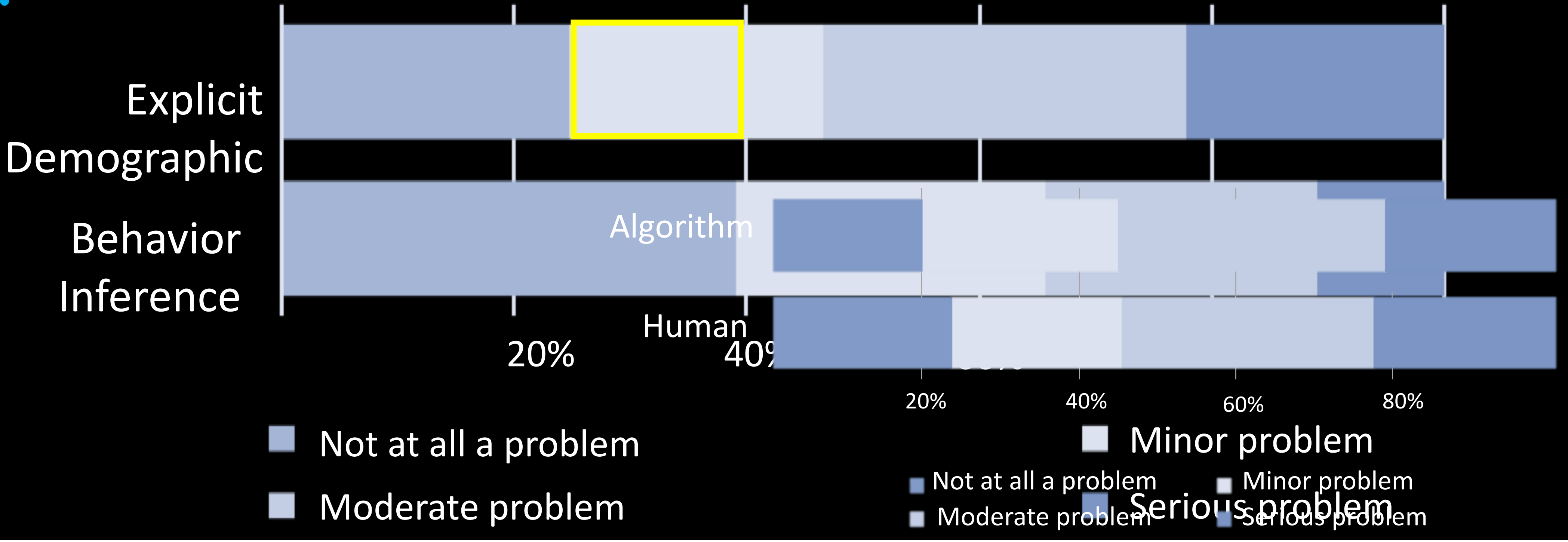
Final Survey & Modeling

Census-representative web panel sample (n=891) with 5-fold CV on trained models

Plane, A., Redmiles, E.M., Mazurek, M.L., and Tschantz, M. *Exploring User Perceptions of Discrimination in Online Targeted Advertising.* **USENIX Security 2017**.

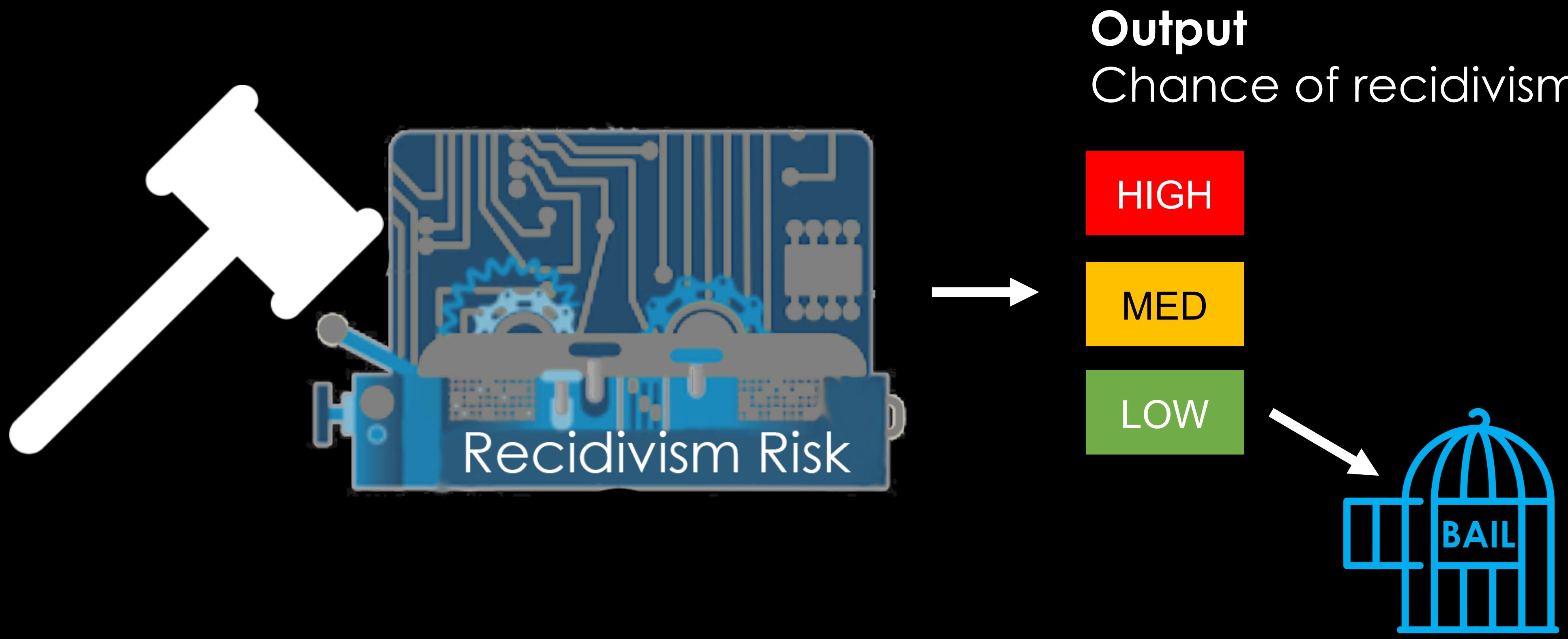
Features are a key factor of perceived fairness

Fairness perception is based on the features (demographic vs. behavior)



Plane, A., Redmiles, E.M., Mazurek, M.L., and Tschantz, M. *Exploring User Perceptions of Discrimination in Online Targeted Advertising.* **USENIX Security 2017**.

COMPAS system helps Florida judges make bail decisions



Output
Chance of recidivism

HIGH

MED

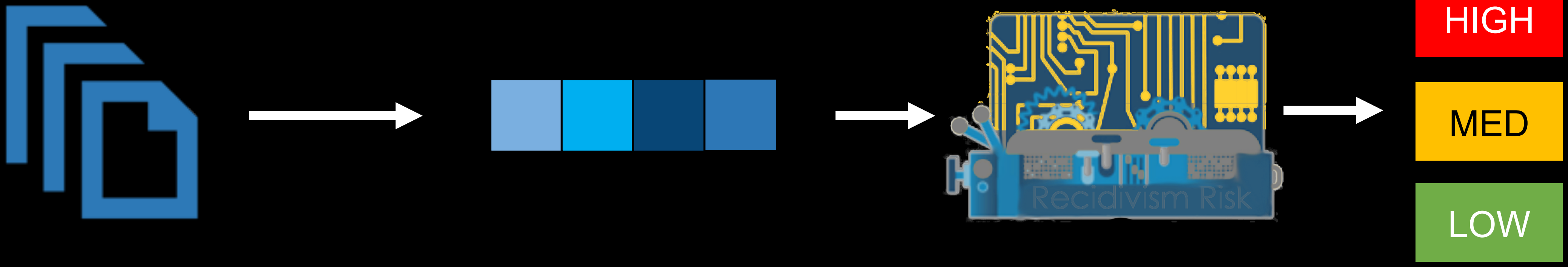
LOW

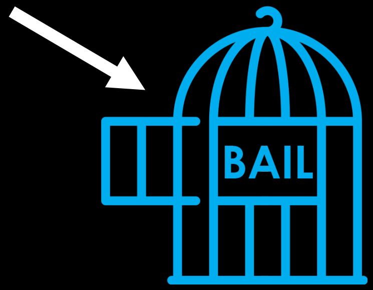
Predict recidivism risk from questionnaire answers

Input
Defendant's answers to COMPAS questionnaire

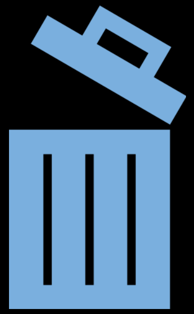
Features
Selected answers to questions

Output
Chance of recidivism





- Current charge
- Criminal history of family and friends
- Performance in School
- Mental health status
- ,,,
- *Nothing Legally Sensitive (Race, Gender, etc.)*



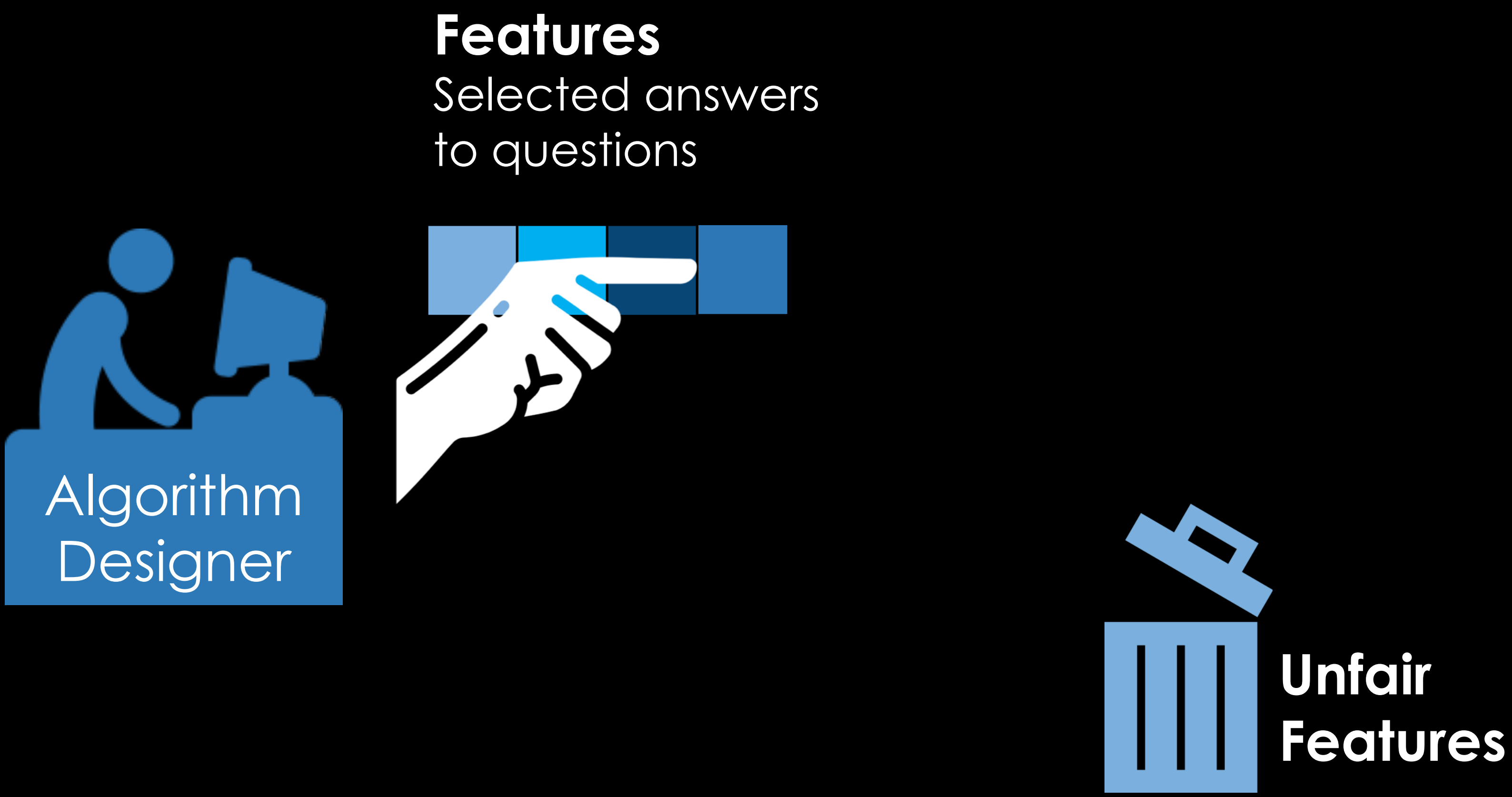
Unfair Features

Analog system: judges admit evidence

Features
Selected answers
to questions

Unfair Features

COMPAS: algorithm designers select features



Features
Selected answers
to questions

Algorithm
Designer

**Unfair
Features**

What If?

We Followed Peoples’ Beliefs About Fairness

Features
Selected answers
to questions

Unfair Features

Survey to assess people's fairness beliefs

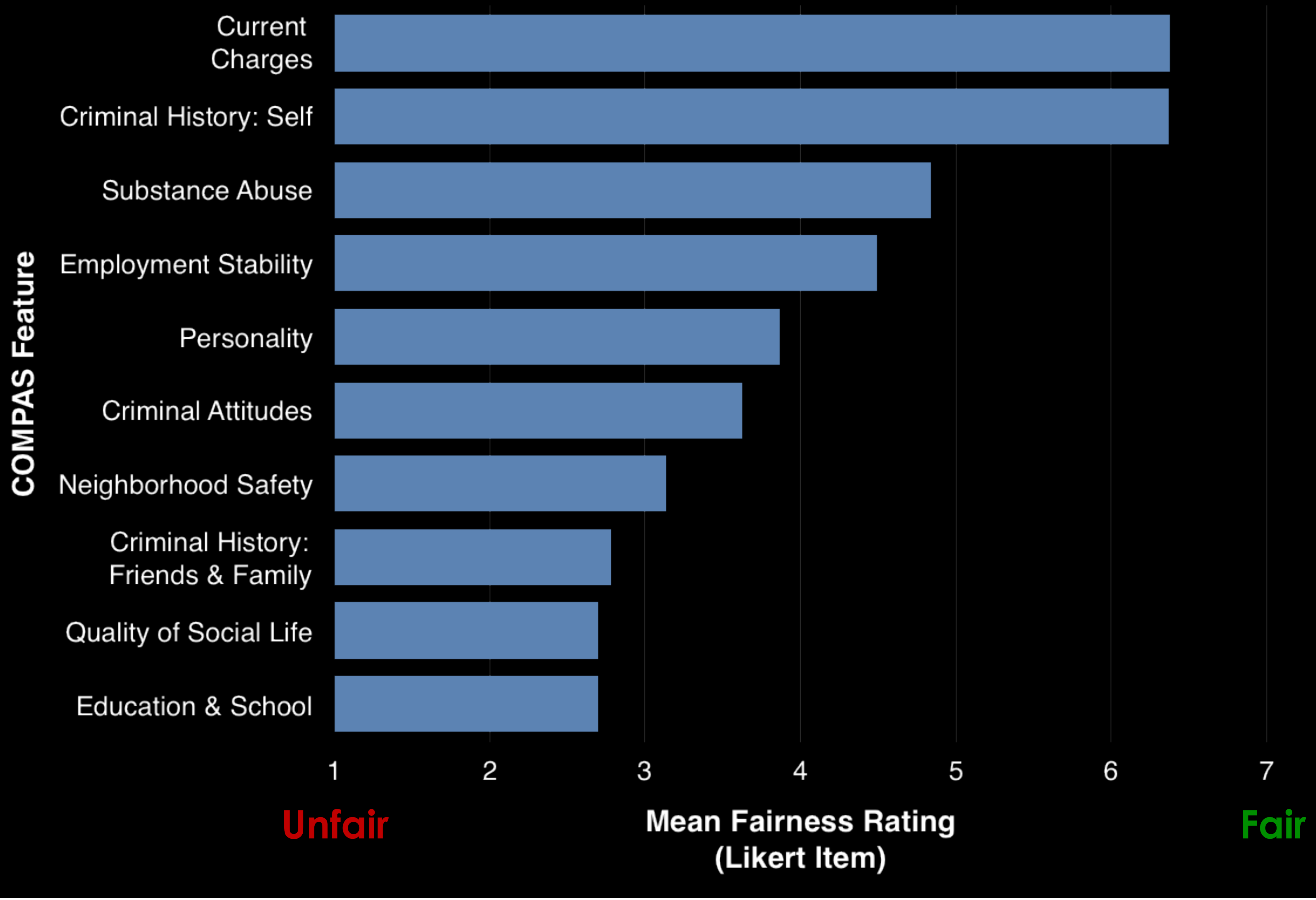
Online survey

Judges in Broward County, Florida, have started using a computer program to help them decide which defendants can be released on bail before trial. The computer program they are using takes into account information about the defendant's **stability of employment and living situation**.

For example, the computer program will take into account the defendant's answer to the following question: **How often do you have trouble paying bills?**

Please rate how much you agree with the following statement:
It is fair to determine if a person can be released on bail using information about their **stability of employment and living situation.**

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., and Weller, A. *Human Perceptions of Fairness in Algorithmic Decision Making.* The Web Conference (**WWW2018**).



Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., and Weller, A. *Human Perceptions of Fairness in Algorithmic Decision Making.* The Web Conference (**WWW2018**).

Lack of consensus on fairness beliefs, why?

How should we set **security** policies?

- **Observe** behavior
- Infer preference
- Make decision

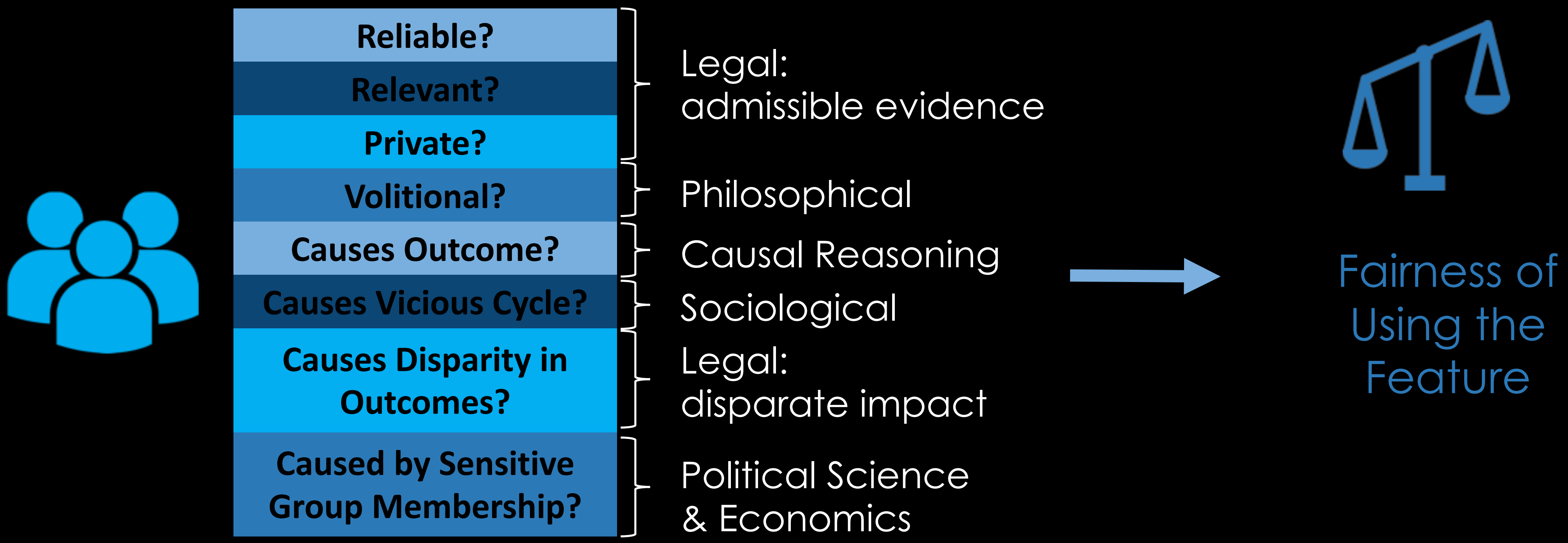
Which features are fair to use in **machine learning**?

- **Ask** preference ✓
- Make decision

What content should be allowed in virtual reality?

- Make decision **together**

People determine ``fairness'' based on eight sub-questions



Reliable?

Relevant?

Private?

Volitional?

Causes Outcome?

Causes Vicious Cycle?

Causes Disparity in Outcomes?

Caused by Sensitive Group Membership?

Legal:
admissible evidence

Philosophical

Causal Reasoning

Sociological

Legal:
disparate impact

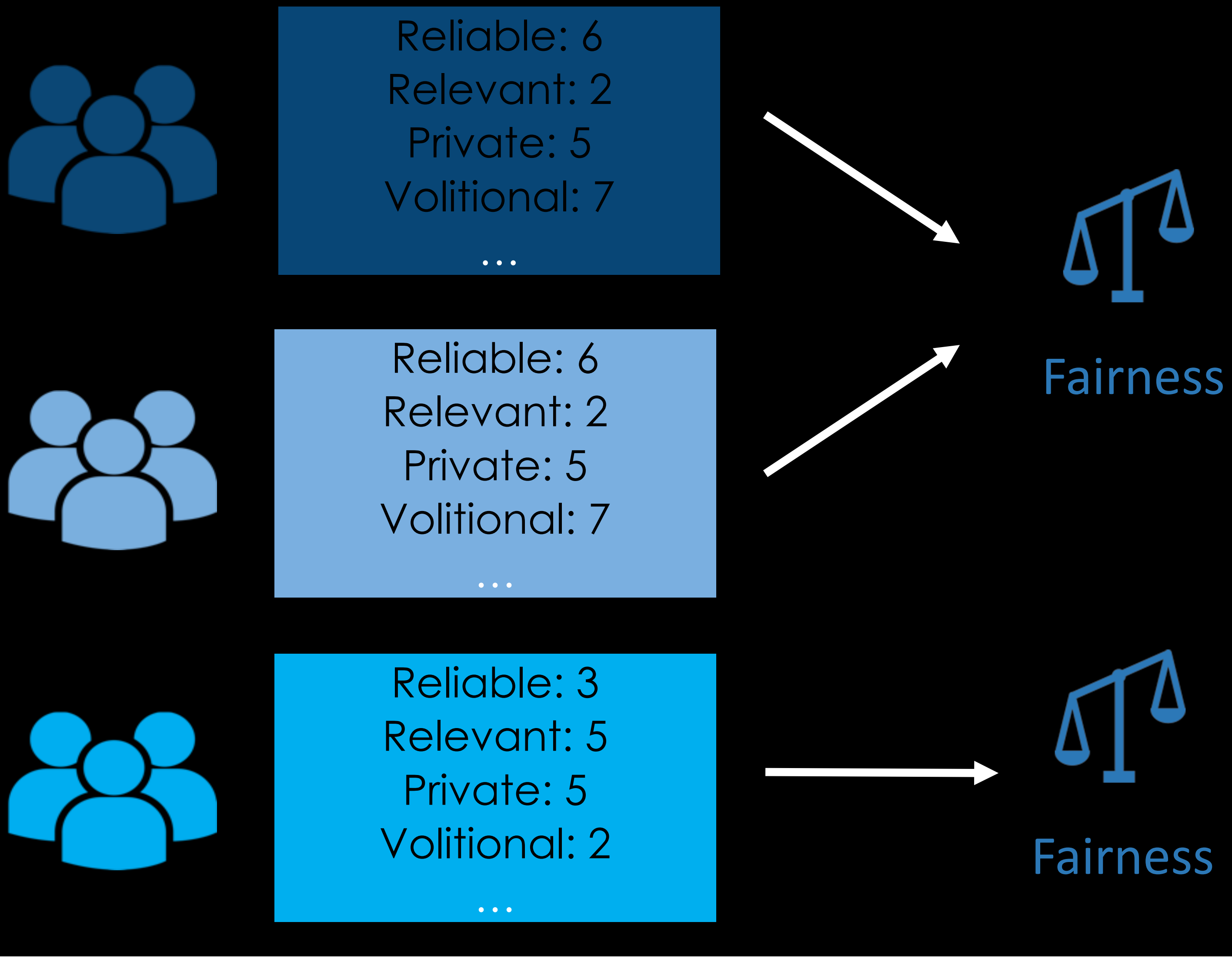
Political Science
& Economics

Fairness of Using the Feature

88% accuracy predicting fairness from property ratings

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., and Weller, A. *Human Perceptions of Fairness in Algorithmic Decision Making.* The Web Conference (**WWW2018**).

Lack of consensus in property ratings, not fairness beliefs



Reliable: 6
Relevant: 2
Private: 5
Volitional: 7
…

Reliable: 6
Relevant: 2
Private: 5
Volitional: 7
…

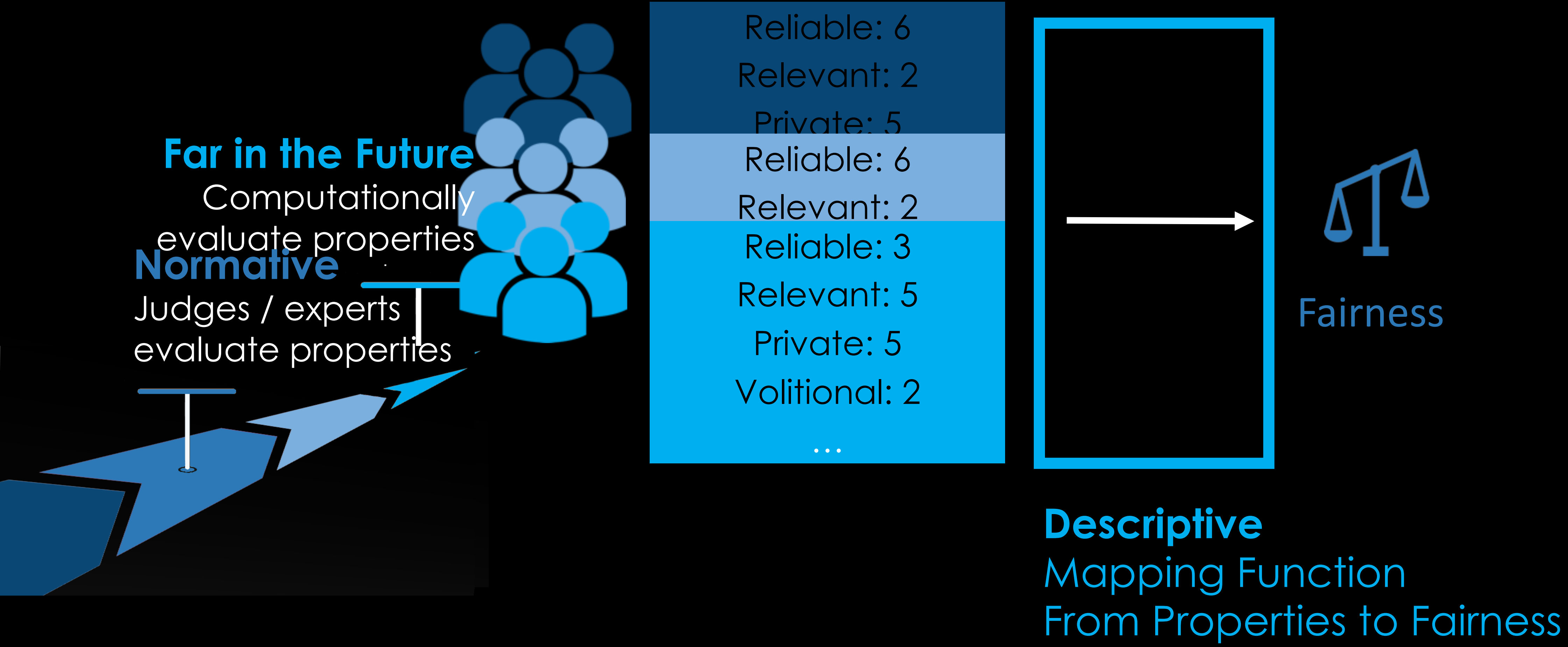
Reliable: 3
Relevant: 5
Private: 5
Volitional: 2
…

Fairness

Fairness

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., and Weller, A. *Human Perceptions of Fairness in Algorithmic Decision Making.* The Web Conference (**WWW2018**).

Descriptive for mapping properties to fairness
Normative to evaluate feature properties



Far in the Future

Computationally evaluate properties

Normative

Judges / experts evaluate properties

Reliable: 6
Relevant: 2
Private: 5

Reliable: 6
Relevant: 2

Reliable: 3
Relevant: 5
Private: 5
Volitional: 2
...

Fairness

Descriptive

Mapping Function
From Properties to Fairness

Grgic-Hlaca, N., Redmiles, E.M., Gummadi, K.P., and Weller, A. *Human Perceptions of Fairness in Algorithmic Decision Making.* The Web Conference (**WWW2018**).

How should we set **security** policies?

- **Observe** behavior
- Infer preference
- Make decision

Which features are fair to use in **machine learning**?

- **Ask** preference ✓
- Make decision ✓

What content should be allowed in **virtual reality**?

- Make decision **together**

Constrain search for features based on fairness threshold

Grgic-Hlaca, N., Zafar, M.B., Gummadi, K. P., Weller, A. AAAI2018.

Can we just make the decision together with the users?

How should we set **security** policies?

Observe behavior

Infer preference

Make decision

Which features are fair to use in **machine learning**?

~~**Ask** preference~~

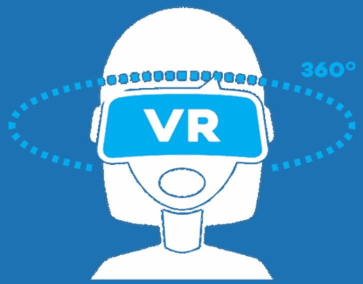
Make decision

What content should be allowed in virtual reality?

Make decision **together**

Virtual Reality

Define guidelines for VR development



Goal

Determine guidelines for for developing VR content

Descriptive Approach

Co-design standards *with* VR developers

Interview Study: VR developers want guidelines

“there’s a quite a big list of unknowns right now in terms of what’s best etiquette for a user and what’s gonna keep the user the most [safe], comfortable, and satisfied”

-- Developer 8

“just the fact of the matter is there are no VR power users. I can count on my hand the number of experienced ‘devs’ I’ve actually met”

-- Developer 5

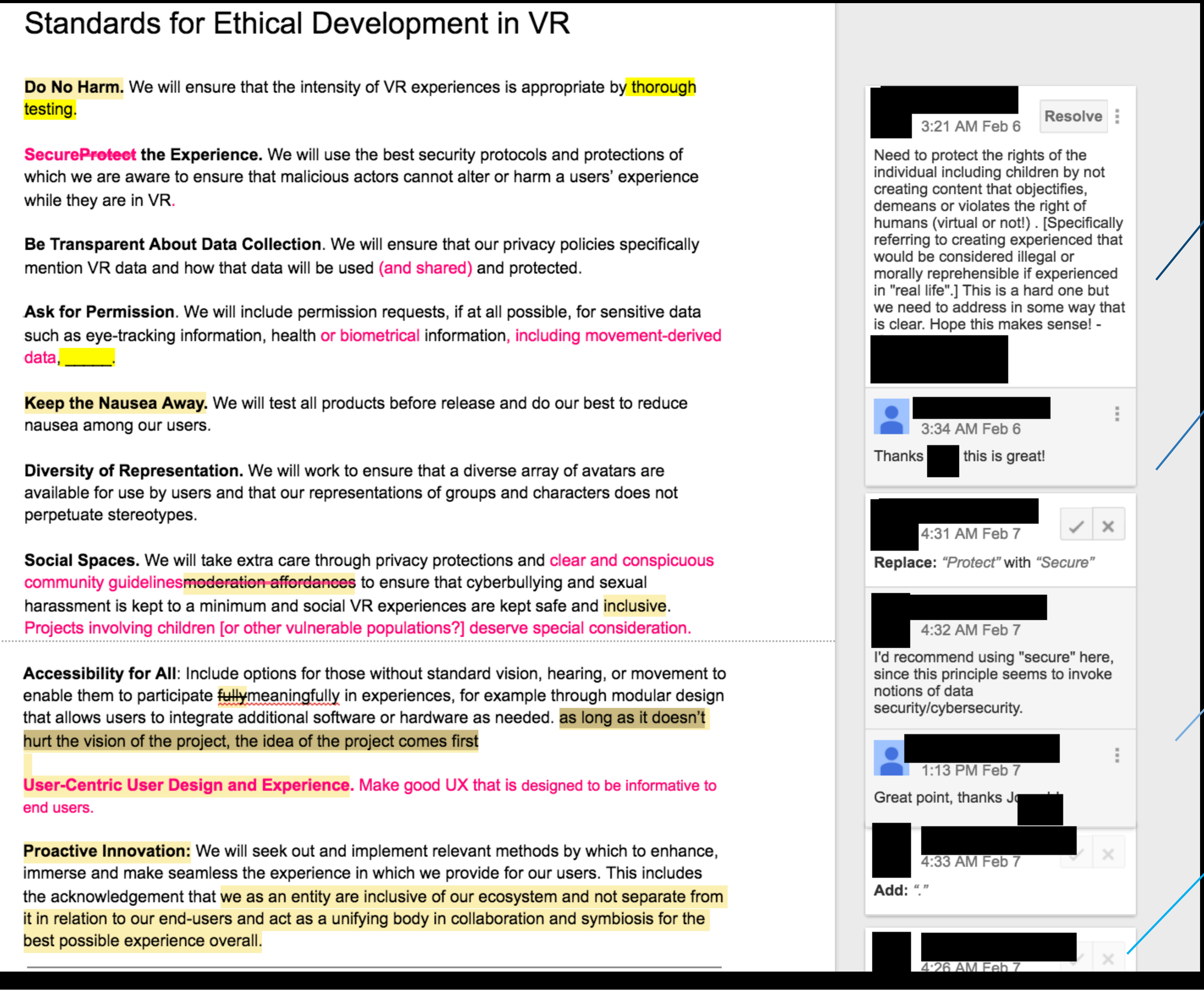
Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**

Code of ethics co-design with developers

Six high level principles drawn by researchers from interview results

Invite 11 online communities of VR developers to edit the draft

Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**



Standards for Ethical Development in VR

Do No Harm. We will ensure that the intensity of VR experiences is appropriate by thorough testing.

Secure~~Protect~~ **the Experience.** We will use the best security protocols and protections of which we are aware to ensure that malicious actors cannot alter or harm a users' experience while they are in VR.

Be Transparent About Data Collection. We will ensure that our privacy policies specifically mention VR data and how that data will be used (and shared) and protected.

Ask for Permission. We will include permission requests, if at all possible, for sensitive data such as eye-tracking information, health or biometrical information, including movement-derived data, ______.

Keep the Nausea Away. We will test all products before release and do our best to reduce nausea among our users.

Diversity of Representation. We will work to ensure that a diverse array of avatars are available for use by users and that our representations of groups and characters does not perpetuate stereotypes.

Social Spaces. We will take extra care through privacy protections and clear and conspicuous community guidelines~~moderation affordances~~ to ensure that cyberbullying and sexual harassment is kept to a minimum and social VR experiences are kept safe and inclusive. Projects involving children [or other vulnerable populations?] deserve special consideration.

Accessibility for All: Include options for those without standard vision, hearing, or movement to enable them to participate ~~fully~~meaningfully in experiences, for example through modular design that allows users to integrate additional software or hardware as needed. as long as it doesn't hurt the vision of the project, the idea of the project comes first

User-Centric User Design and Experience. Make good UX that is designed to be informative to end users.

Proactive Innovation: We will seek out and implement relevant methods by which to enhance, immerse and make seamless the experience in which we provide for our users. This includes the acknowledgement that we as an entity are inclusive of our ecosystem and not separate from it in relation to our end-users and act as a unifying body in collaboration and symbiosis for the best possible experience overall.

[redacted]
3:21 AM Feb 6
Resolve

Need to protect the rights of the individual including children by not creating content that objectifies, demeans or violates the right of humans (virtual or not!) . [Specifically referring to creating experienced that would be considered illegal or morally reprehensible if experienced in "real life".] This is a hard one but we need to address in some way that is clear. Hope this makes sense! -

[redacted]

[redacted]
3:34 AM Feb 6

Thanks [redacted] this is great!

[redacted]
4:31 AM Feb 7
✓ ✕

Replace: *"Protect"* with *"Secure"*

[redacted]
4:32 AM Feb 7

I'd recommend using "secure" here, since this principle seems to invoke notions of data security/cybersecurity.

[redacted]
1:13 PM Feb 7

Great point, thanks J[redacted]

[redacted]
4:33 AM Feb 7
✓ ✕

Add: "."

[redacted]
4:26 AM Feb 7
✓ ✕

1053 Views

245 potential editors

Engagement equiv. to Wikipedia editing (10%)

19 editors

7 sharers

40 contributions

Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**

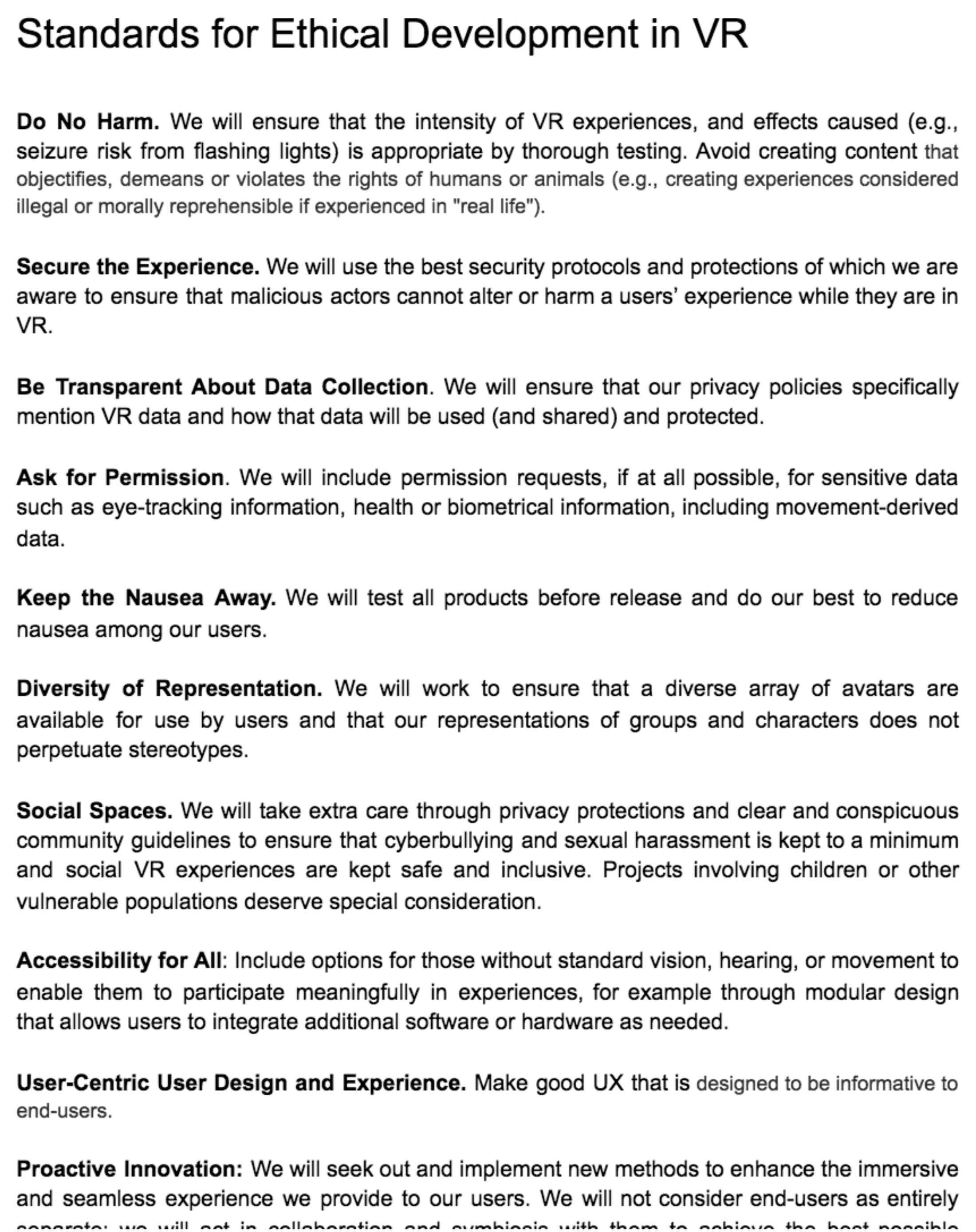
Developers reached consensus on 10 principles

Six high level principles drawn by researchers from interview results

Invite 11 online communities of VR developers to edit the draft

Trace ethnographic analysis of editing process [see paper]

Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**



Standards for Ethical Development in VR

Do No Harm. We will ensure that the intensity of VR experiences, and effects caused (e.g., seizure risk from flashing lights) is appropriate by thorough testing. Avoid creating content that objectifies, demeans or violates the rights of humans or animals (e.g., creating experiences considered illegal or morally reprehensible if experienced in "real life").

Secure the Experience. We will use the best security protocols and protections of which we are aware to ensure that malicious actors cannot alter or harm a users' experience while they are in VR.

Be Transparent About Data Collection. We will ensure that our privacy policies specifically mention VR data and how that data will be used (and shared) and protected.

Ask for Permission. We will include permission requests, if at all possible, for sensitive data such as eye-tracking information, health or biometrical information, including movement-derived data.

Keep the Nausea Away. We will test all products before release and do our best to reduce nausea among our users.

Diversity of Representation. We will work to ensure that a diverse array of avatars are available for use by users and that our representations of groups and characters does not perpetuate stereotypes.

Social Spaces. We will take extra care through privacy protections and clear and conspicuous community guidelines to ensure that cyberbullying and sexual harassment is kept to a minimum and social VR experiences are kept safe and inclusive. Projects involving children or other vulnerable populations deserve special consideration.

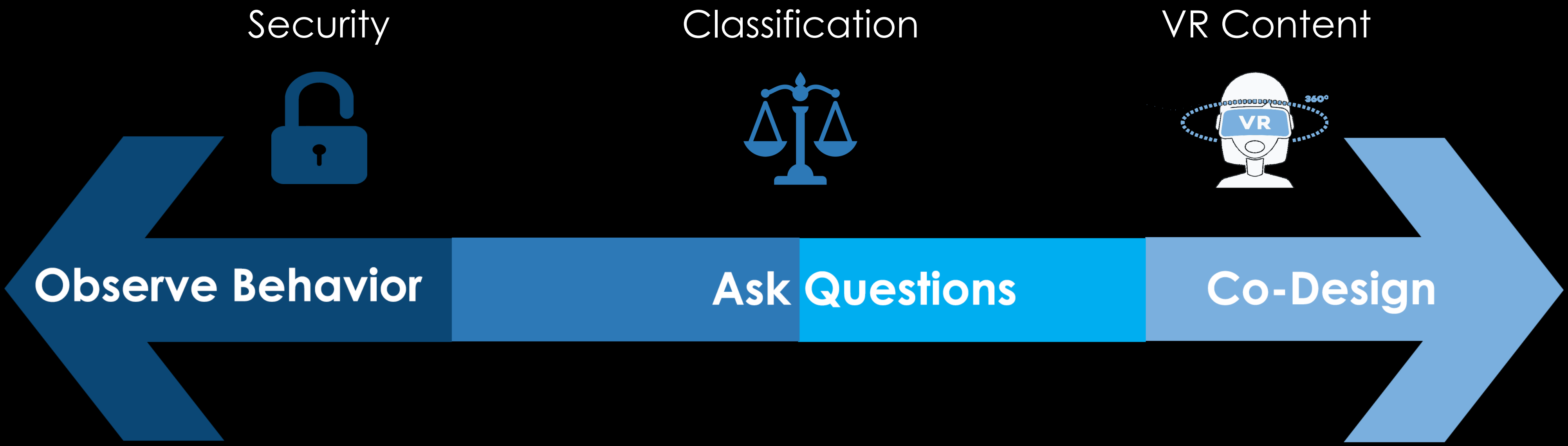
Accessibility for All: Include options for those without standard vision, hearing, or movement to enable them to participate meaningfully in experiences, for example through modular design that allows users to integrate additional software or hardware as needed.

User-Centric User Design and Experience. Make good UX that is designed to be informative to end-users.

Proactive Innovation: We will seek out and implement new methods to enhance the immersive and seamless experience we provide to our users. We will not consider end-users as entirely separate; we will act in collaboration and symbiosis with them to achieve the best possible

Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**

Different methods are appropriate for different problems



Security

Classification

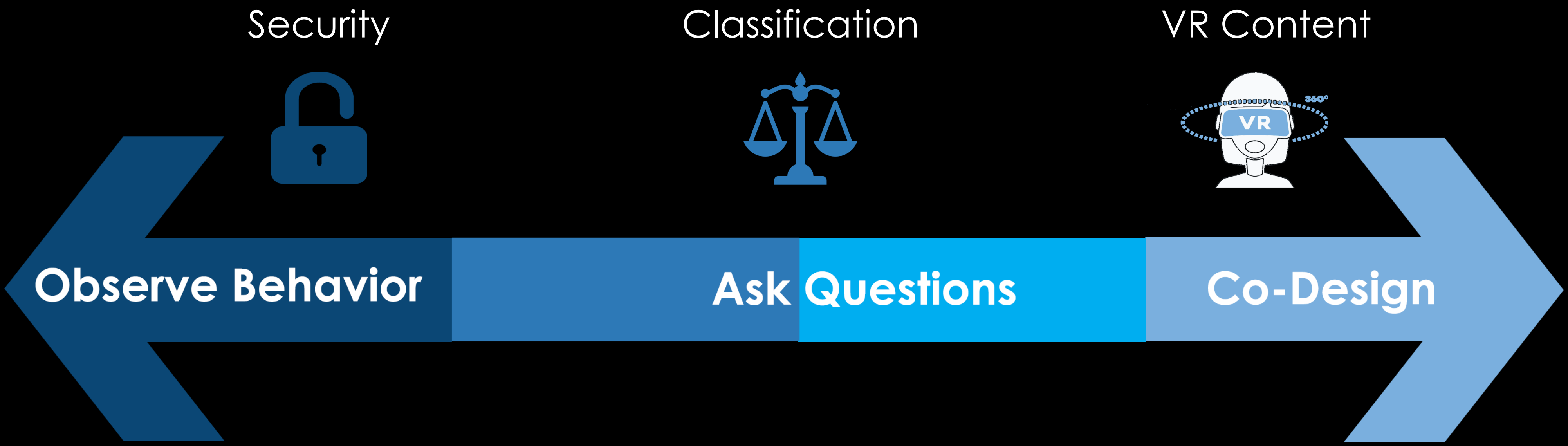
VR Content

Observe Behavior

Ask Questions

Co-Design

Methods have prerequisites:
observation and question-asking require consistency



Security

Classification

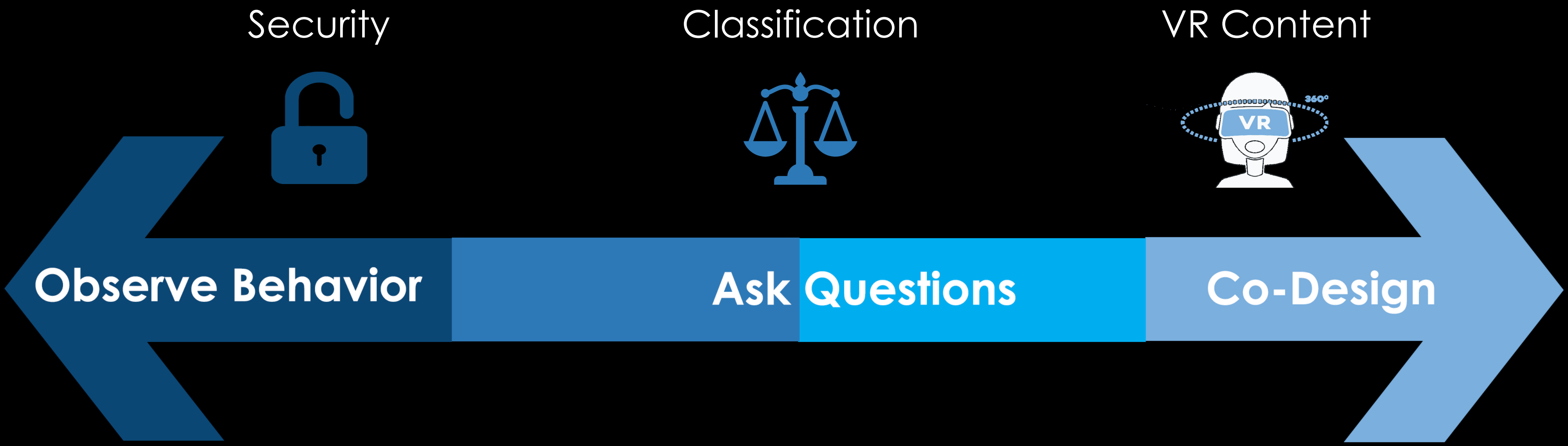
VR Content

Observe Behavior

Ask Questions

Co-Design

Co-design requires recruiting users you think will make "good" choices or A LOT of users



Why Not Have VR Users Co-Design, Too?

Researchers normatively decided that small group of users with homogenous, exclusive opinions weren't good first-round participants

…If you use VR, most likely you [also use] Reddit because there's a certain type of crowd that's really into this, you know?

"somebody who has a lot of money and has a premium setup you know...I mean you are talking people with 4 plus sensors.

I'll be more concerned about virtual crimes and bullying once VR becomes more accessible to the "general public."

YOU CAN'T SIT WITH US

Users

Adams, D., Bah, A., Barwulor, C., Musabay, N., Pitkin, K., and Redmiles, E.M. *Ethics Emerging: the Story of Privacy and Security Perceptions in Virtual Reality.* **SOUPS2018.**

Descriptive vs. Normative: always a balance

Security

Normative expert effectiveness judgement

- Future: compute effectiveness

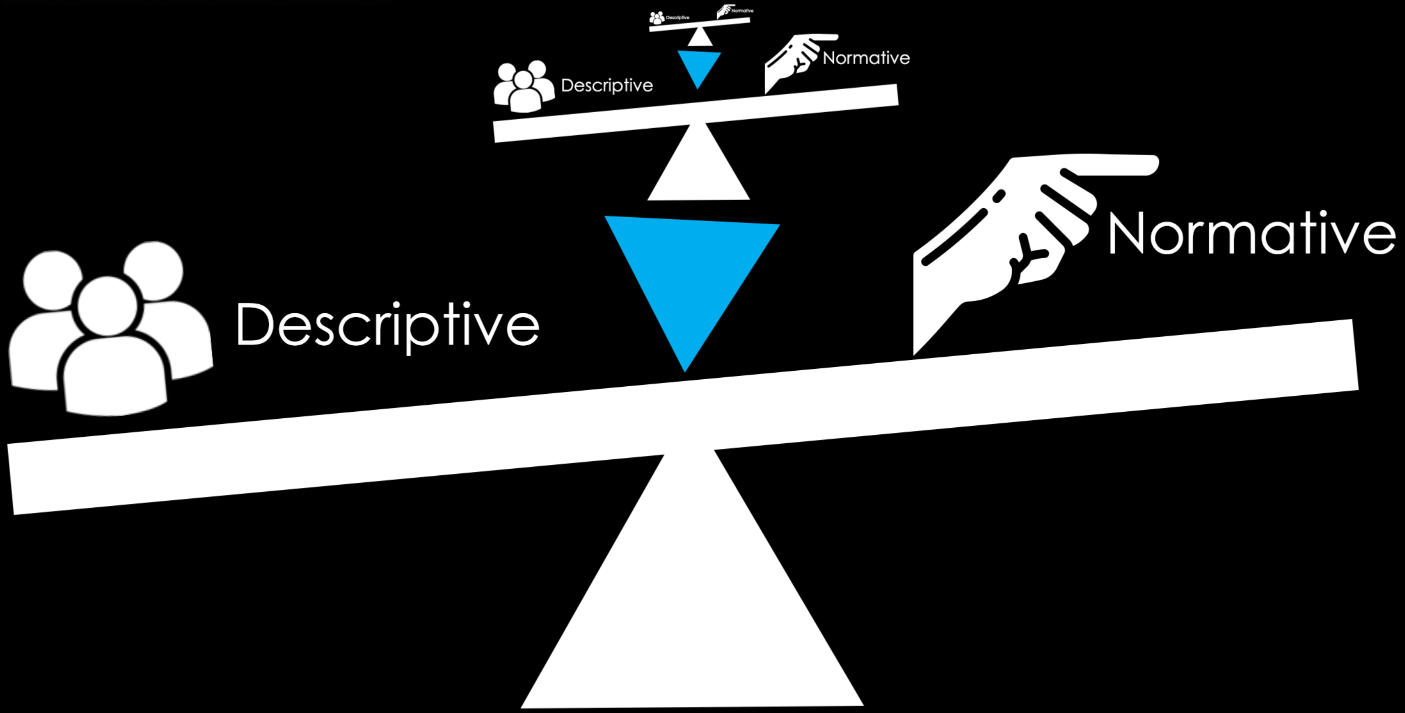
Machine Learning

Normative expert property judgements

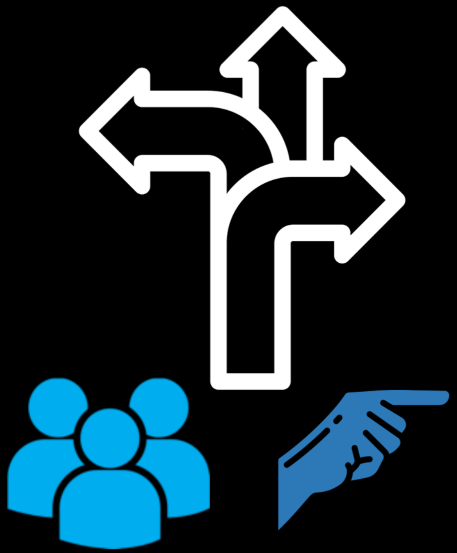
- Future: compute property values

Virtual Reality

Normative researcher judgement of *who* to include in descriptive approach

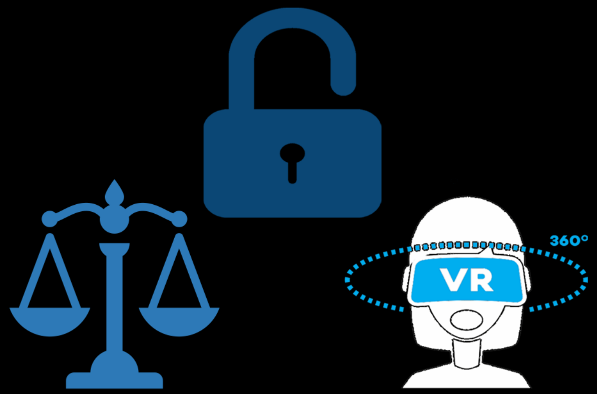


At what are the humans best?
the experts
the computing systems

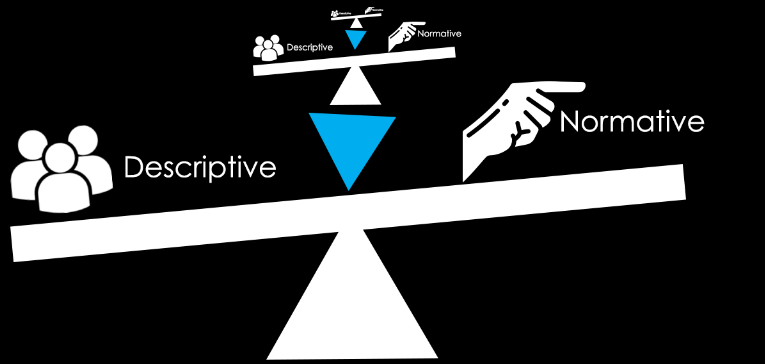


Explore descriptive solutions to computational problems:

learning best practices from people's preferences / behavior



Through examples in security, machine learning, and virtual reality



Illustrate how different balances between normative & descriptive could be achieved

Learning from the People

From Normative to Descriptive Solutions
to Problems in Security, Privacy & Machine Learning

Elissa M. Redmiles
@eredmil1
eredmiles@gmail.com