Human-Centered Privacy in Machine Learning

Bailey Kacsmar

PUPS Research Lab

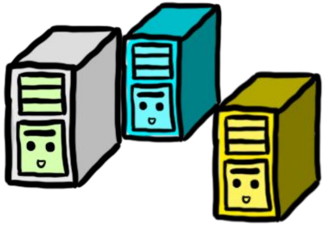
UNIVERSITY OF ALBERTA

amii

The Plan

- Challenges of ensuring security and privacy for ML
- How we design and evaluate security and privacy for ML
- Why “Human-Centered”

Why Privacy and ML?



A company wants to analyze data

But the data has privacy implications for the data subjects

Researchers develop technical solutions

Beyond Data

Google and Mastercard Cut a Secret Ad Deal to Track Retail Sales

Google found the perfect way to link online ads to store purchases: credit card data

By Mark Bergen and Jennifer Surane
August 30, 2018, 3:43 PM EDT *Updated on August 31, 2018, 12:40 PM EDT*

washingtonpost.com

Now for sale: Data on your mental health

Drew Harwell

These retailers share customer data with Facebook's owner. Customers may not have been told | CBC News

Thomas Daigle · CBC News · Posted: Feb 07, 2023 4:00 AM EST | Last

Home Depot didn't get customer consent before sharing data with Facebook's owner, privacy watchdog finds | CBC News

Catharine Tunney · CBC News · Posted: Jan 26, 2023 9:53 AM
Updated: January 27

Double-double tracking: How Tim Hortons knows where you sleep, work and vacation

James McLeod June 15, 2020 In : Canada Privacy 0 1,169 11 min read

Beyond Data as a Concept

Google and Mastercard ...
Deal to Tra...

Google found the ...
card data

By Mark Bergen and Jennife...
August 30, 2018, 3:43 PM ED...

ADOBE / CREATORS / TECH

Adobe's new terms of service aren't the problem – it's the trust

/ The reaction from Adobe's customers to a small update highlights the growing lack of faith surrounding big tech companies and their AI tools.

Creatives are fearful of how Adobe's adoption of generative AI will impact their privacy and rights over their work. Illustration by Haein Jeong / The Verge

By Jess Weatherbed, a news writer focused on creative industries, computing, and internet culture. Jess started her career at TechRadar, covering news and hardware reviews.

Jun 7, 2024, 1:37 PM MDT

11 Comments (11 New)

If you buy something from a Verge link, Vox Media may earn a commission. See our ethics statement.

...ental

Home Depot ...
consent befo...
Facebook's o...
finds | CBC Ne...

Catharine Tunney · CBC Ne...
Updated: January 27

...and Vacation

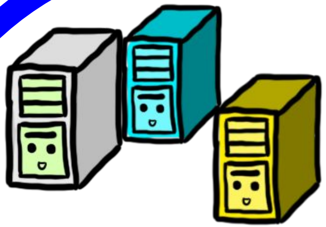
...e 15, 2020 In : Canada Privacy 0 1,169 11 min read

Understanding the Challenge:

Privacy and ML?

What makes this hard? What's the risk?

Why Privacy and ML?



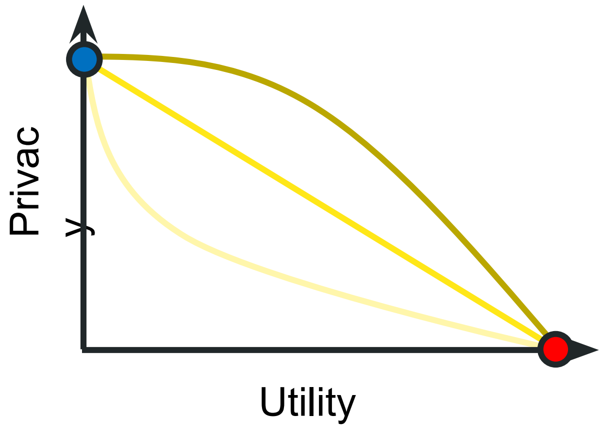
A company wants to analyze data

But the data has privacy implications for the data subjects

Researchers develop technical solutions

The Privacy-Utility trade-off

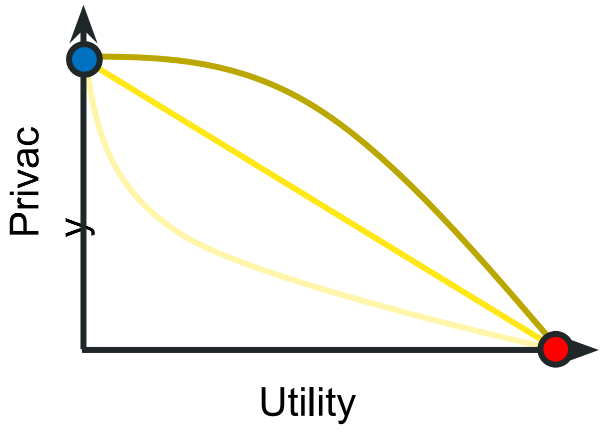
- Given any metric for privacy and for utility, they are usually at odds:



- **Q:** How do you design a system that provides maximum utility?
- **Q:** How do you design a system that provides maximum privacy?
- Designing a system that provides a good privacy-utility trade-off is hard!

The Privacy-Utility trade-off

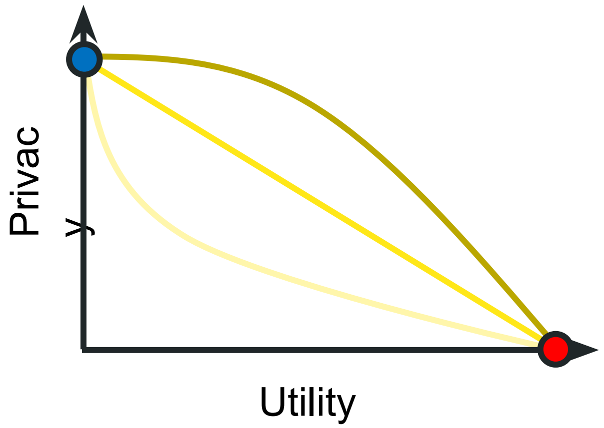
- Given any metric for privacy and for utility, they are usually at odds:



- How do you design a system that provides maximum utility?
  - You design it without privacy in mind
- How do you design a system that provides maximum privacy?
  - ..?
- Designing a system that provides a good privacy-utility trade-off is hard!

The Privacy-Utility trade-off

- Given any metric for privacy and for utility, they are usually at odds:

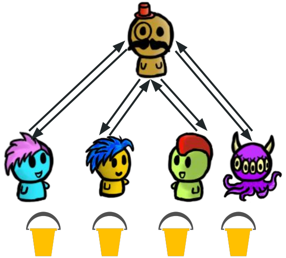


- How do you design a system that provides maximum utility?
  - You design it without privacy in mind
- How do you design a system that provides maximum privacy?
  - You don't design it
- Designing a system that provides a good privacy-utility trade-off is hard!

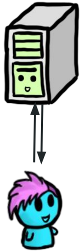
Private Computation

Balancing Privacy and Utility

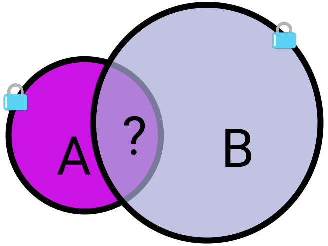
Private Computations



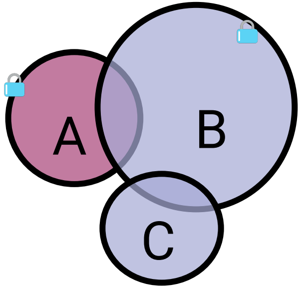
Private Machine Learning



Private Query Processing



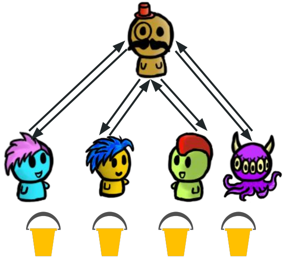
Private Set Intersection



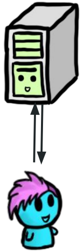
Multiparty Computations

Private Computations Class

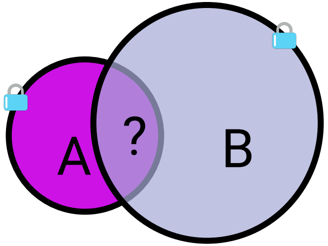
Define, **what** is being protected, from **whom**,
and under what **conditions** this protection will hold.



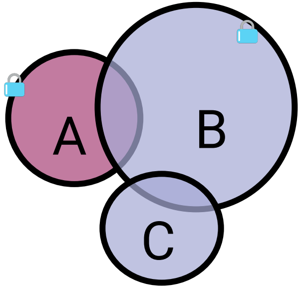
Private Machine Learning



Private Query Processing



Private Set Intersection



Multiparty Computations

Technical Guarantees Types

- Statistical
- Computational
- Information Theoretical

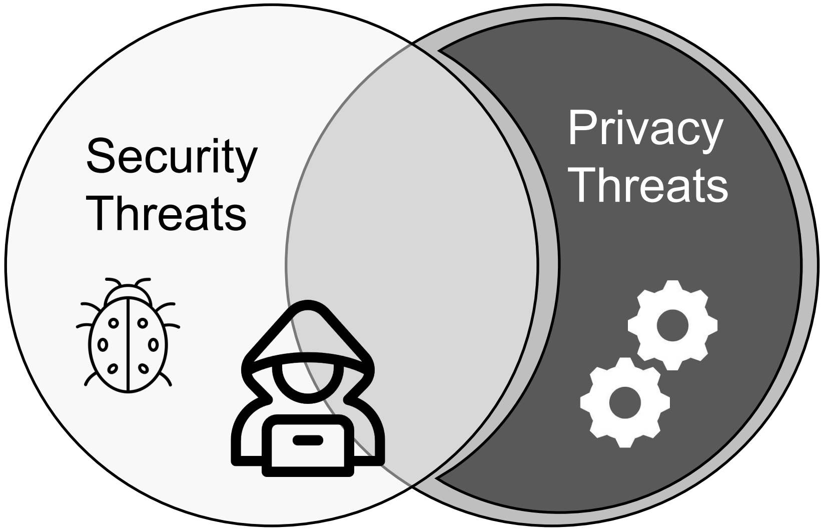
Quantifying Privacy: Theoretical Notions

- **Syntactic** notions of privacy: these are computed on the leaked or released data. They are data dependent
    - K-anonymity, l-diversity, t-closeness, etc
- **Semantic** notions of privacy: these are computed on the data release mechanism itself, and they hold regardless of the data (data independent)
    - Mostly Differential Privacy

Quantifying Privacy: Empirical Notions

- The performance of an **inference attack** e.g., the attacker error, accuracy, true positive rate, false positive rate, etc
- Can provide an **upper bound** on privacy

Quantifying Privacy versus Security



Westin's (1967)

An entity's **ability to control** how, when, and to what extent personal information about it is communicated to others

For privacy, focus on the **harms** (consequences) caused by privacy violations.

Harms from Privacy Violations

Financial

Physical

Targeted Ads

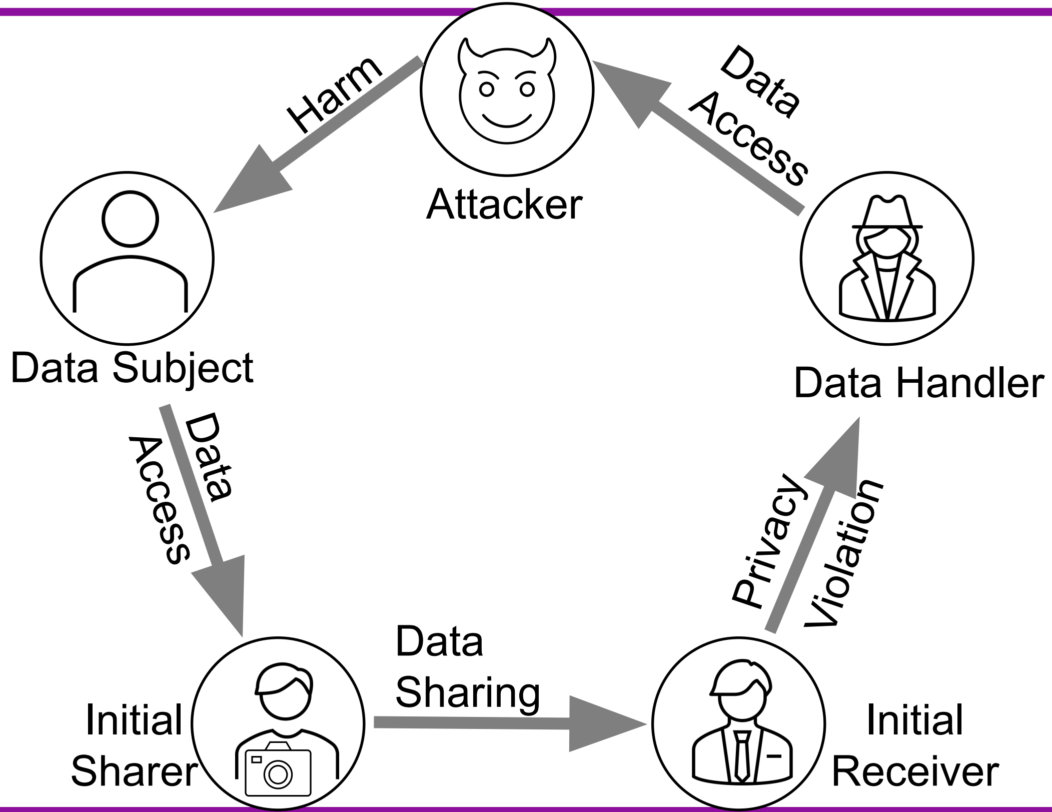
Social

Legal Prosecution

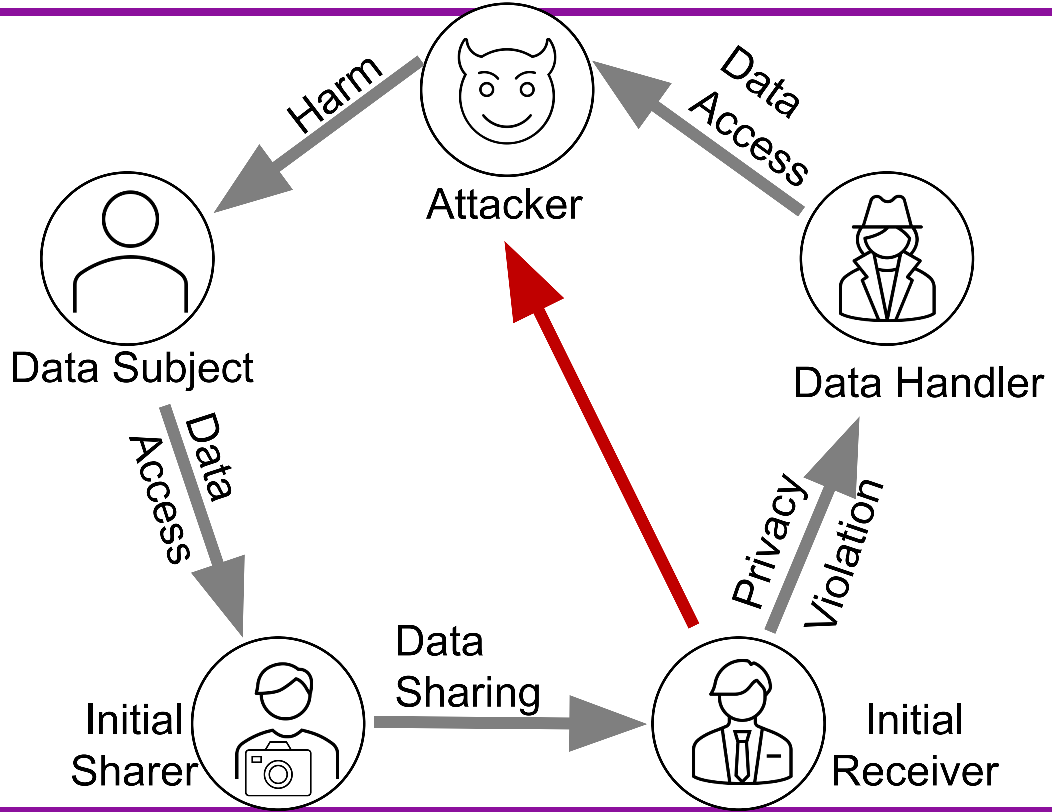
Mental

Mass Surveillance

Privacy Violation Life-Cycle



Privacy Violation Life-Cycle



Adversarial Thinking

- Think like an adversary to understand the ***vulnerabilities*** of a system and develop ***protection techniques***.
- When designing inference attacks, we also apply **Kerckhoff's principle** (or Shannon's maxim), adapted to privacy

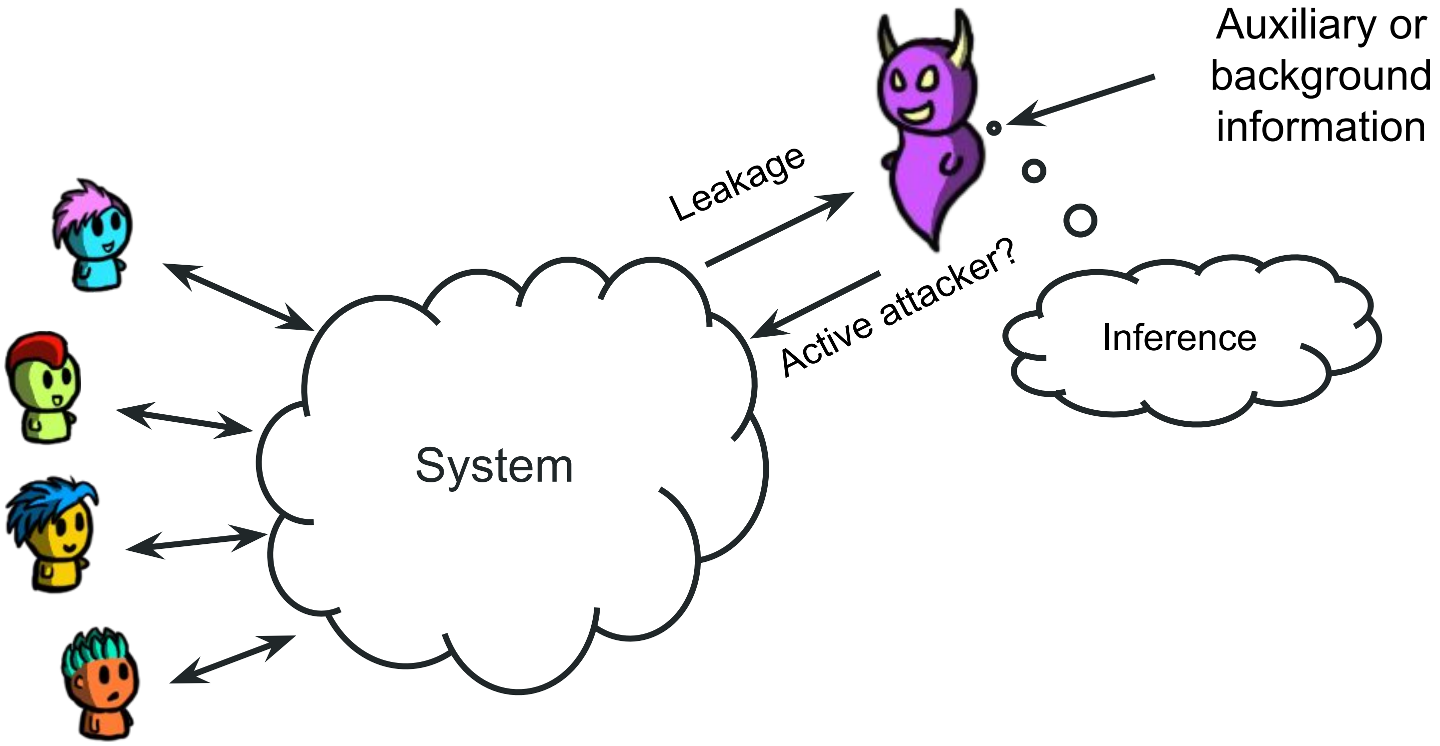
Adversarial Thinking

- Think like an adversary to understand the ***vulnerabilities*** of a system and develop ***protection techniques***.
- When designing inference attacks, we also apply **Kerckhoff's principle** (or Shannon's maxim), adapted to privacy

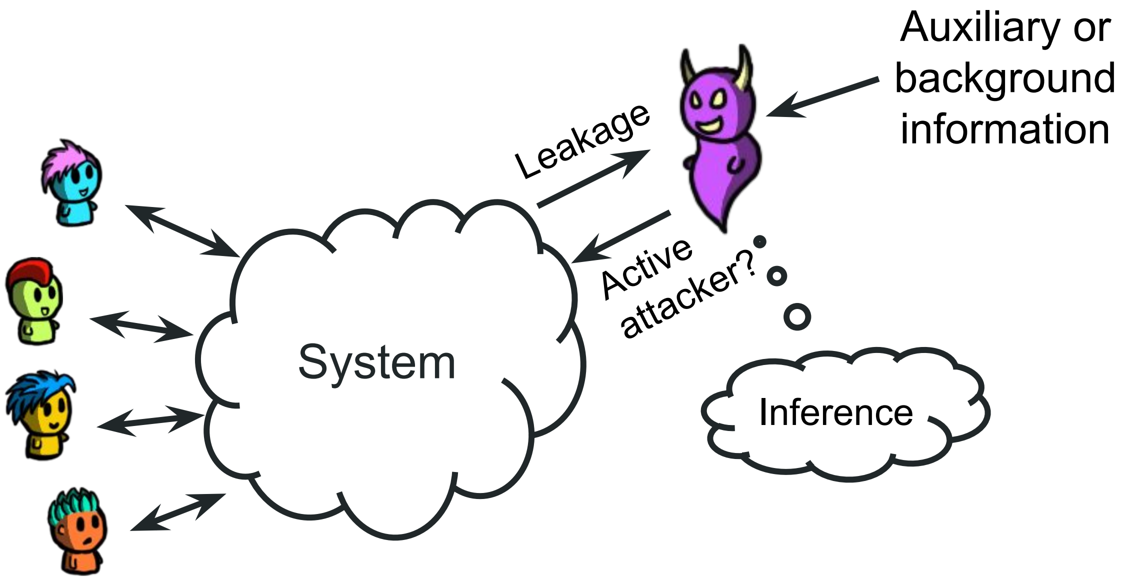
Assume the adversary knows how the system works

- There are **no hidden parameters** other than the users' data
- The adversary can **even know some rough distribution**

What are inference attacks?

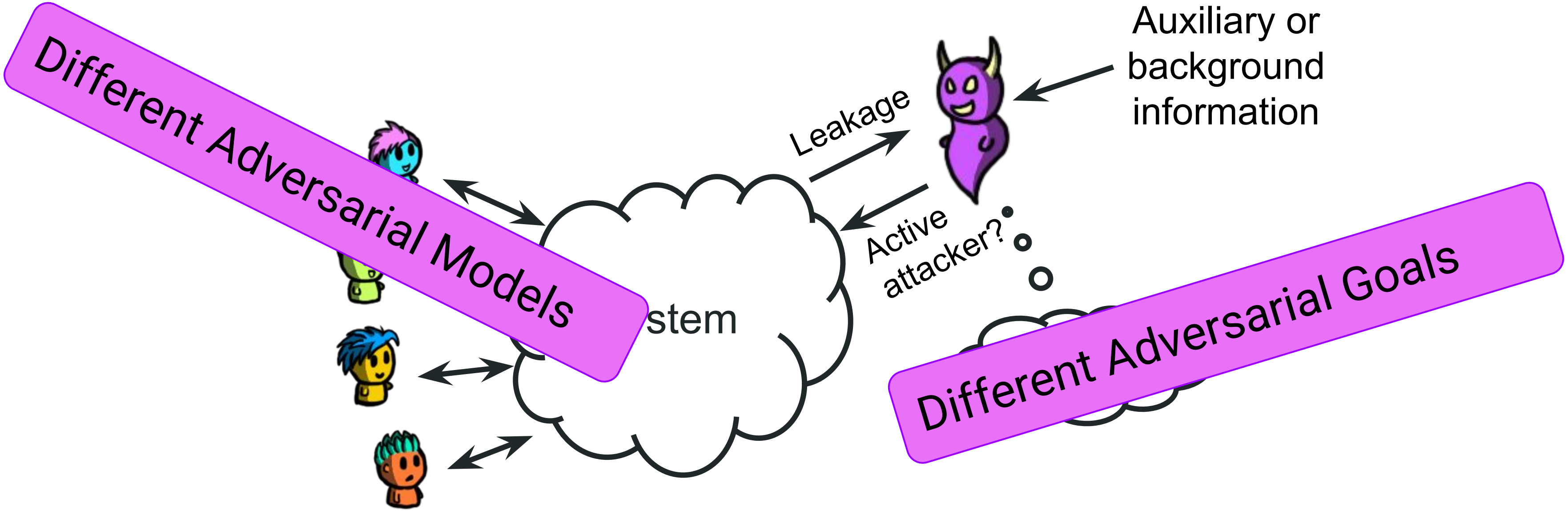


What are inference attacks?



Goal: Learn something (non-trivial) and privacy sensitive from the system

What are inference attacks?



Goal: Learn something (non-trivial) and privacy sensitive from the system

Inference Attacks: Goals and Abilities

- **Goals**:
  - Infer data
  - Infer a property of the data
  - Infer the presence (membership) of some data
  - Infer the behavior of a user
  - Infer some attributes of a data sample
  - Infer dependencies among the data
  - …

Inference Attacks: Goals and Abilities

- **Goals**:
  - Infer data
  - Infer a property of the data
  - Infer the presence (membership) of some data
  - Infer the behavior of a user
  - Infer some attributes of a data sample
  - Infer dependencies among the data
  - …
- **Abilities**:
  - Statistical tools (estimation theory, detection theory, maximum likelihood, Bayesian inference…)
  - Combinatorics
  - Heuristics
  - Machine learning
  - …

Designing a System Aware of Inference Attacks

For any system that relies on users' data, there are two goals:

- **Utility:** Design a system that provides benefits to its users and the service provider
- **Privacy:** Design a system that provides protection against inference attacks

Q: What are "utility" and "privacy"? How do we "measure" them?

Designing a System Aware of Inference Attacks

For any system that relies on users' data, there are two goals:

- **Utility:** Design a system that provides benefits to its users and the service provider
- **Privacy:** Design a system that provides protection against inference attacks

Q: What are "utility" and "privacy"? How do we "measu

It's complicated...

Building Blocks for Private Machine Learning:

What are we protecting and how?

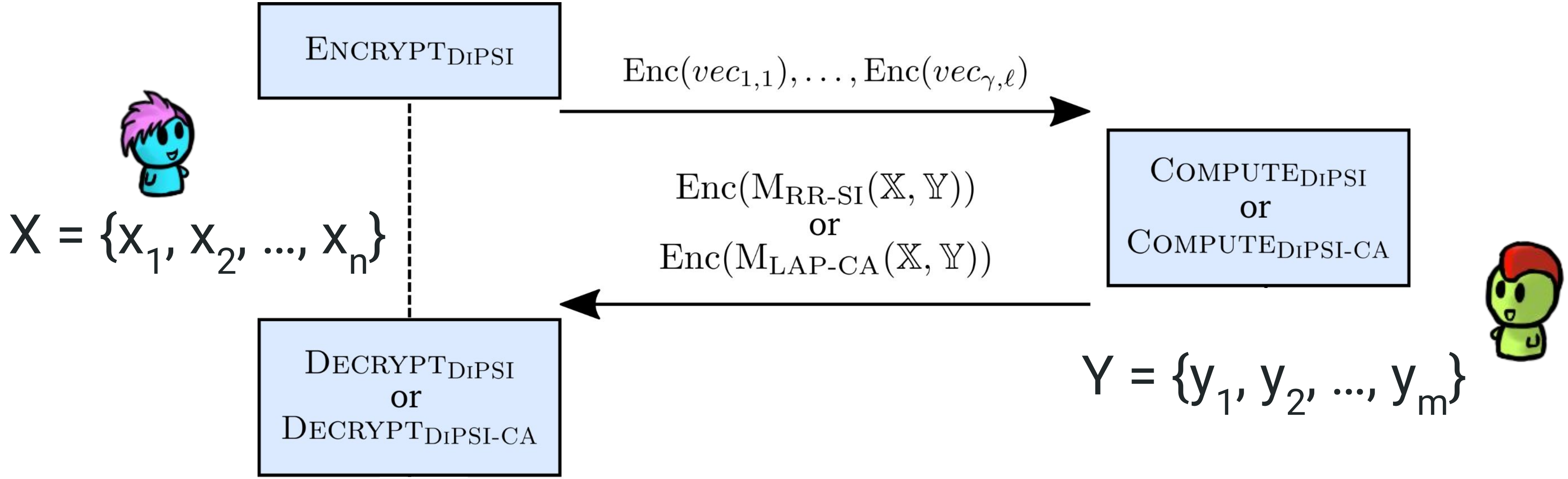
A Private Computation? Cryptography!

I want to learn
$Z = X \cap Y$

$X = \{x_1, x_2, \ldots, x_n\}$

$Y = \{y_1, y_2, \ldots, y_m\}$

Private Set Intersection



Kacsmar Khurram, Lukas, Norton, et al. "Differentially private two-party set operations." In 2020 IEEE European Symposium on Security and Privacy (EuroS&P), pp. 390-404. IEEE, 2020.

Private Computation and Machine Learning?

Training Data	Models	Inferences/Outputs

Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Private Computation and Machine Learning?

Training Data	Models	Inferences/Outputs
Unintentional Leakage		Intentional Leakage

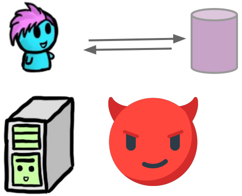
Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Private Computation and Machine Learning?

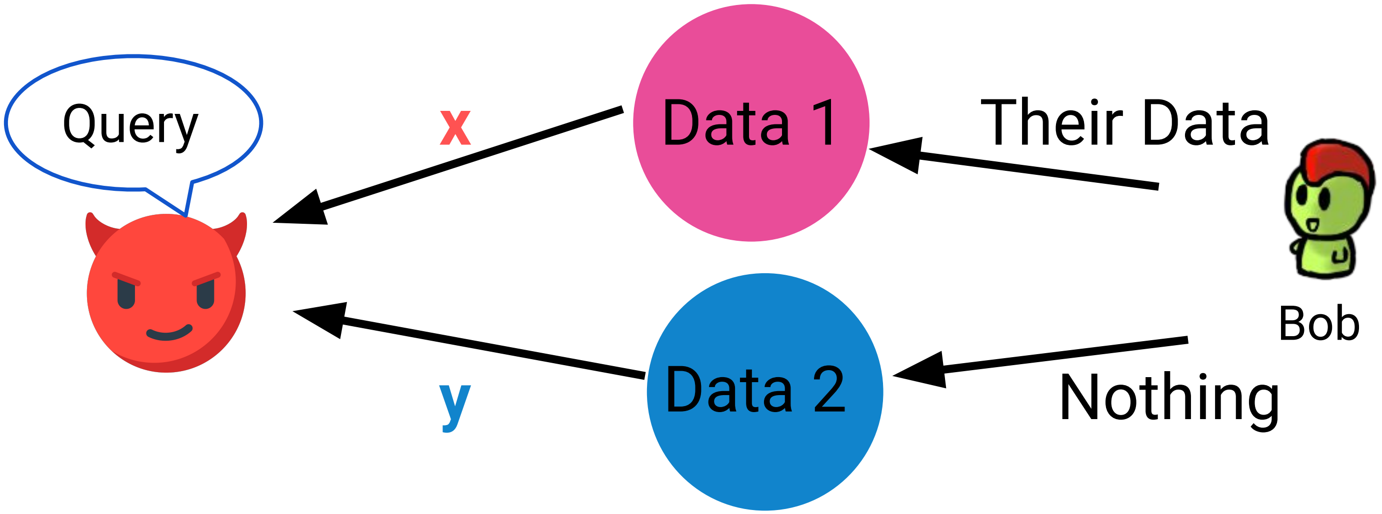
Training Data	**Models**	**Inferences/Outputs**
Unintentional Leakage		**Intentional Leakage**
Data Subject	**Data Owner**	**Access Control**

Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Technical Privacy: Differential Privacy Intuition



Technical Privacy



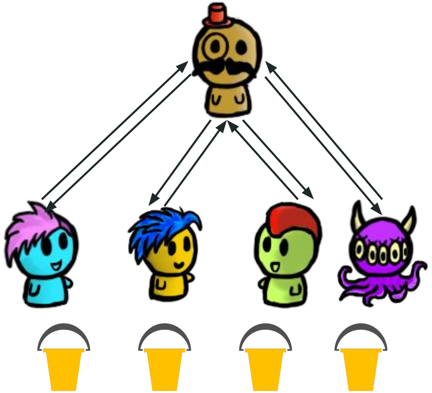
Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Differential Privacy and Machine Learning

- DP-SGD
- Individualized Differential Privacy (PATE)
- More…

However, still **require expertise** for deployment

Distribution of Trust



Federated Learning
PLUS something

- Distribution alone is not private
- SMPC is…expensive
- But…

Not putting all the eggs in one basket, will always have appeal.

Challenge:

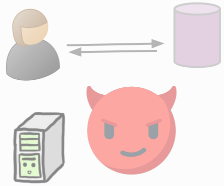
Is it enough? What about the other vectors…

Challenge:

Is it enough? What about the other vectors…

Consent and Communication

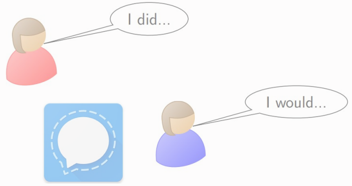
A Wider View of Technical Privacy



Technical Privacy

Conceptual Privacy

Legal Privacy

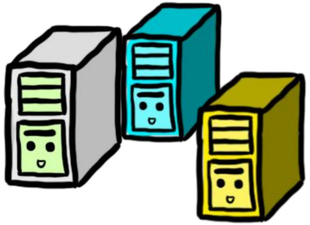


Usable Privacy

Understanding privacy notions and behaviours, **right to privacy**, and privacy expectations

M. Oates, et al. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration." Proceedings on Privacy Enhancing Technologies 2018.

Why Private Computation?



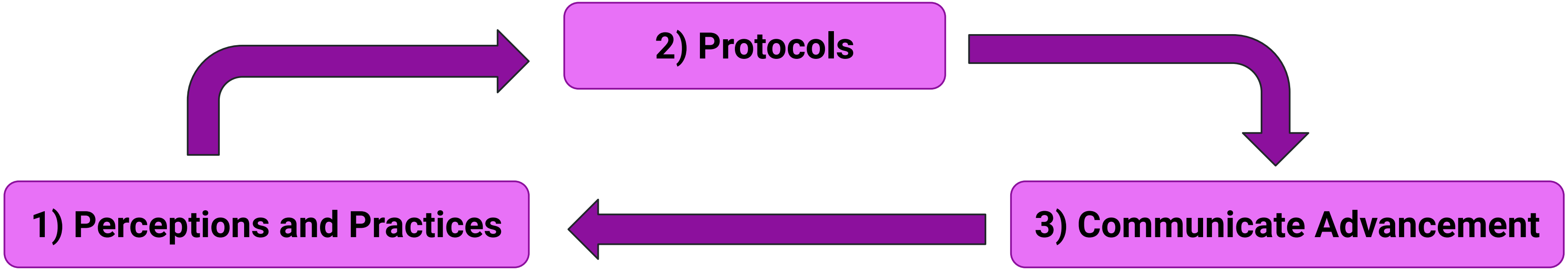
A company wants to analyze data

But the data has privacy implications for the data subjects

Researchers develop technical solutions

In what ways does private computation matter to people?

Human-Centered Design



"...that aims to make systems usable and useful by **focusing on the users, their needs and requirements**, ... counteracts possible adverse effects of use..." - ISO 9241-210:2019(E)

Implications of Sharing Structures

- Disambiguate Third Parties

> PetSmart's privacy policy states: "We may share the information we collect with companies that provide support services to us."

- Current systems contains insufficient information to support preferences impacted by sharing type
- Privacy preferences fluctuate with any change to context
- Number of parties, trusted parties, purpose, etc. all influence acceptability, regardless of technical privacy

Kacsmar, Tilbury, Mazmudar, Kerschbaum. Caring about Sharing: User Perceptions of Multiparty Data Sharing. *USENIX Security 2022*

Perceptions and Expectations

- What do data subjects understand?
- How is a data subject’s willingness to share impacted?
- How do data subjects perceive the risks?

What they “want”

What they “need”

Build towards those attributes

Kacsmar, Duddu, Tilbury, Ur, and Kerschbaum. Comprehension from Chaos: Towards Informed Consent for Private Computation. *2023 ACM SIGSAC Conference on Computer and Communications Security* (CCS).

The Scenarios

Wage Equity	Census Analysis
Ad Conversion	Contact Discovery

Contact Discovery Conceptual Example

The app wants to **determine the common contacts** between the new user and the existing users via…

1. …the new user shares all their contact information with the social media app.
2. … the new user shares **a modified version** of their contact information…**such that** the social media app does not learn non-users…thus, **this means**…

The Interview

Expectations and Term Awareness → Initial Definition and Baseline → Scenario Assessment

↓

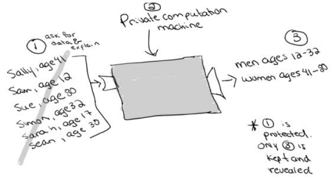
Collective Explanation Activity ← General Perceptions ← Inference Attack Perceptions

Participant Comprehension and Expectations

First Attempt

Second Attempt

Secure computation is a way that a company analyzes your data. The final analysis will be made public [at access location]. However, your specific data is protected and cannot be traced back to you nor can your specific data points be traced back to you. The analysis will be specifically [example], and this is being done because [purpose].



This is the information we're getting from you, but, rest assured, only Part Three will be shown. You can trust us to keep your information private. <If true>This information will only be used for this project and nothing else in the future.

Final Consensus

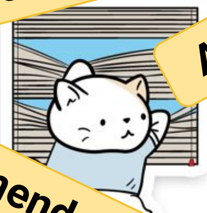
Participant Comprehension and Expectations

Brief

Overlapping

Unsuccessful

First Attempt

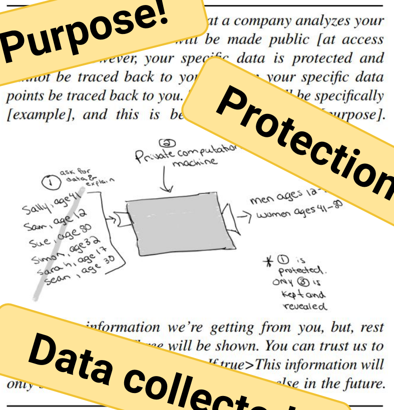


Descriptive!

Accuracy!

Recommendations!

Second Attempt



Purpose!

Protections!

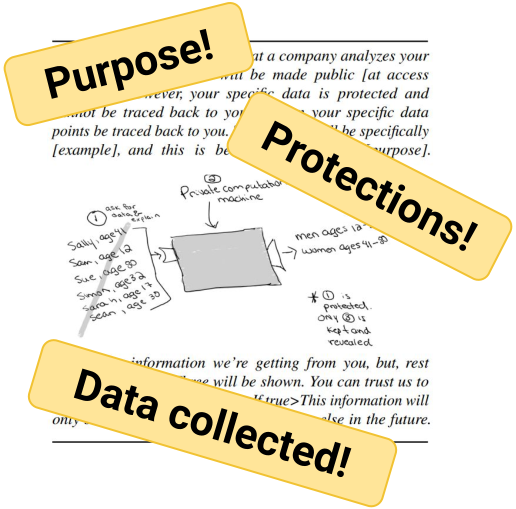
Data collected!

Final Explanation

Participant Comprehension and Expectations

First Attempt

Second Attempt



Final Explanation

Unconcerned with details of the mechanism, **impact** matters

Impact of Private Computation

"…they're trying to make it sound a little bit better" (P19).

"…it feels a little bit more protected that way" (P12)

Bounded Impact of Private Computation

Intentions Matter

Divulge the Details

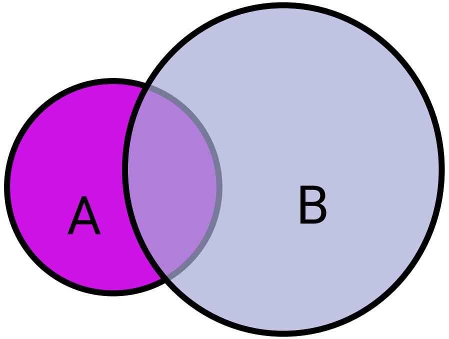
Regulate the Restrictions

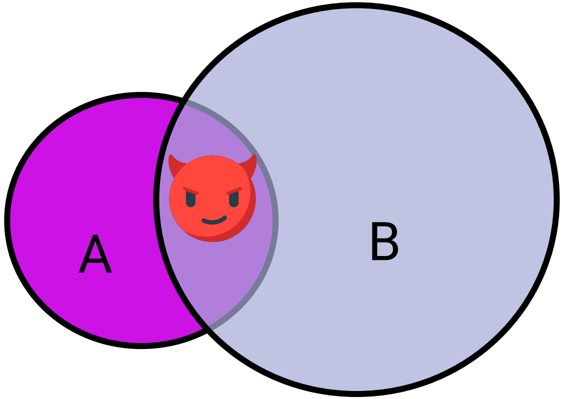
Consent Above All

"At the end of the day, they're still like learning specific things about me" (P7)

Awareness of Unique Threat Models

Alice





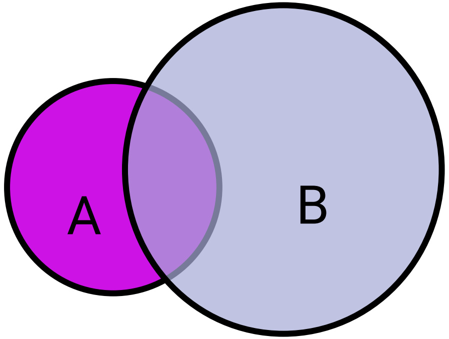
Joins Social App

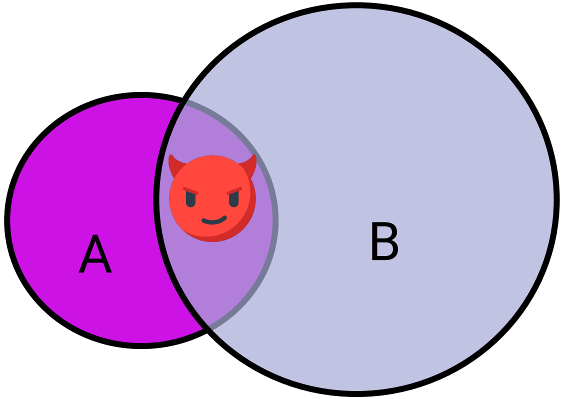
Contact Discovery

Real Identity Connected

Awareness of Unique Threat Models

Alice





Joins Social App

Contact Discovery

Real Identity Connected

There exist, and will continue to exist risks that cannot be regulated by technology

Takeaways

- **Protections provided by protocols** and constructions **do not encompass the full range of risks** experienced **by individuals** in society
- Privacy mitigation techniques are a treatment and not a cure for data privacy concerns
- People find private computation plausible, but they **care about the context, not the math**

Takeaways

- **Protections provided by protocols** and constructions **do not encompass the full range of risks** experienced **by individuals** in society
- Privacy mitigation techniques are a treatment and not a cure for data privacy concerns
- People find private computation plausible, but they **care about the context, not the math**

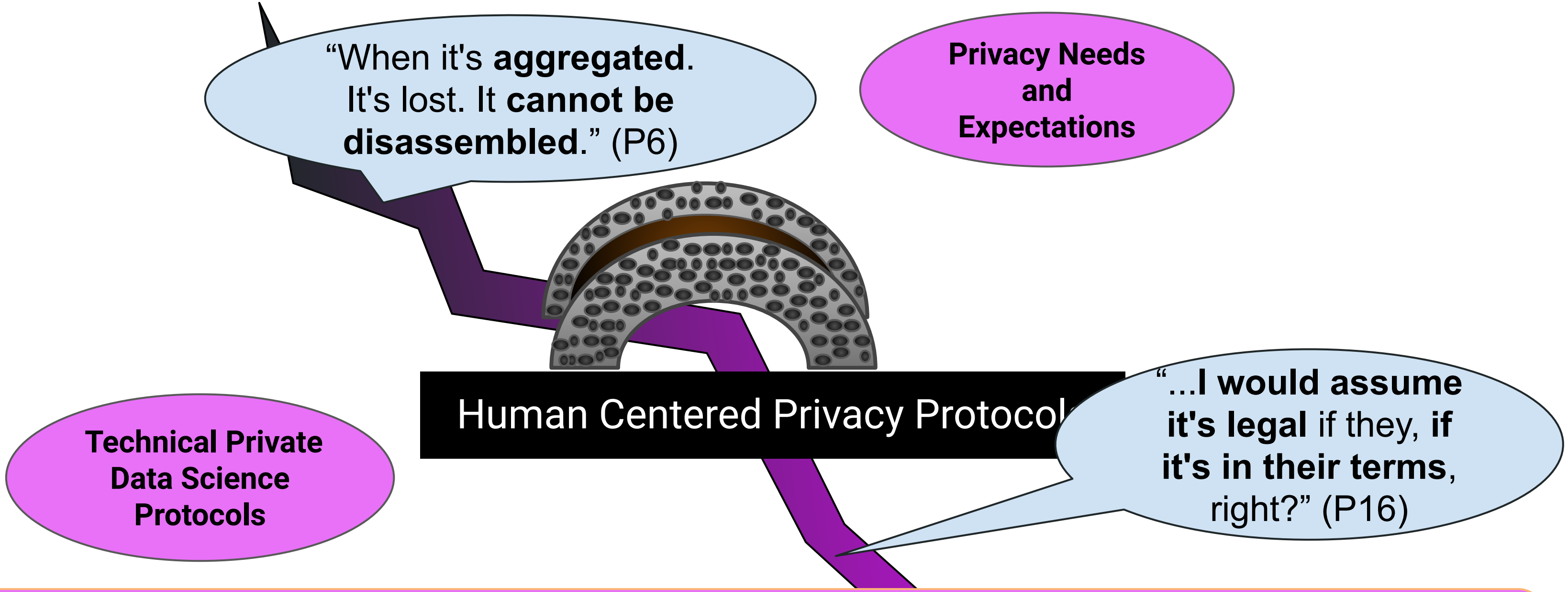
People can reason about private computation; let them.

Thanks!

Bonus Slides

Caring about sharing…

Human Centered Technical Privacy Solutions



Goal: Determine how to best develop technical protocols such that they provide meaningful privacy guarantees to the subjects of the data.

Towards Privacy by Design, Core Tenets

- User centric
- Embedding privacy into the design
- Having privacy as the default configuration
- Ensuring privacy across the whole software life-cycle

Build out Structures for North America

- How do companies share data?
- Who do they share it with?
- Who are the companies?
- When do they share it?
- What do they share?

The Canadian tech company that changed its mind about using your tax return to sell stuff | CBC Radio

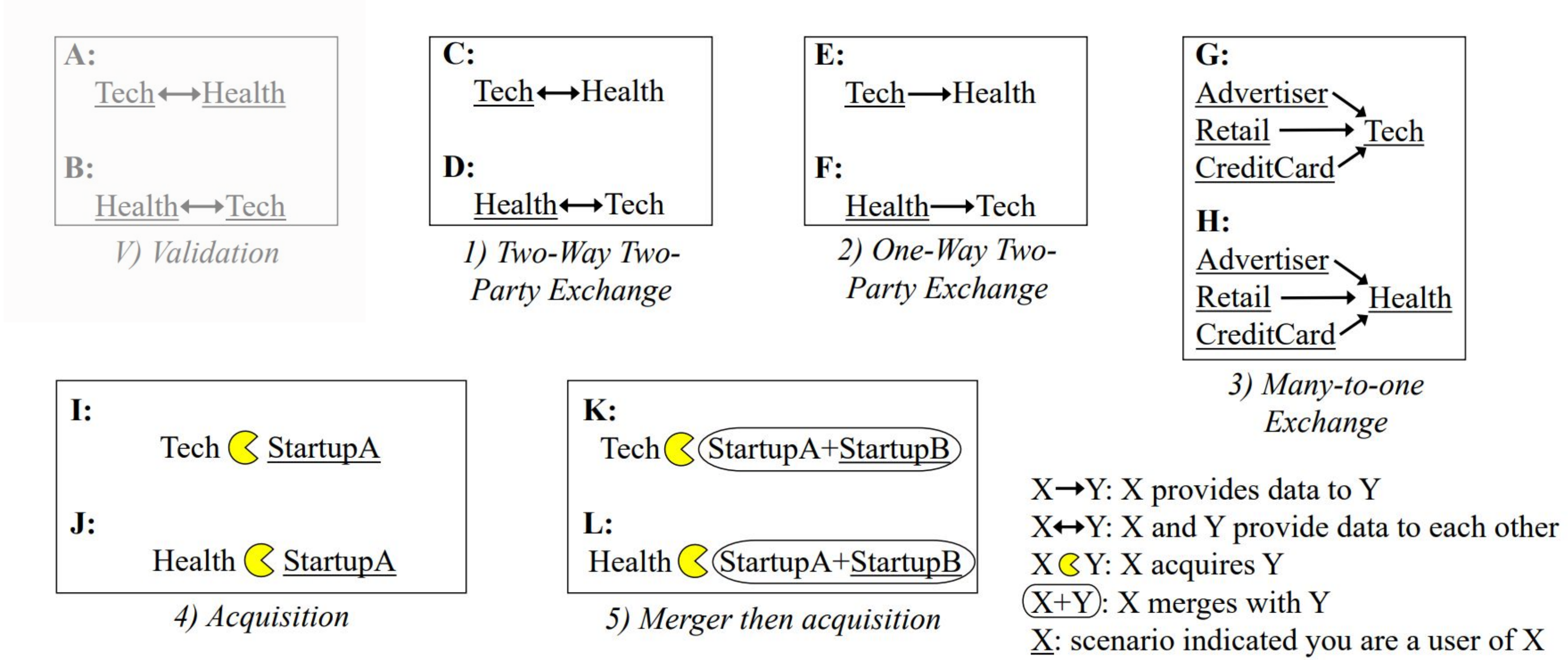
CBC Radio · Posted: Feb 23, 2020 4:00 AM EST | Last Updated: February 23, 2020

What happens to your data when a company dies? - The Parallax

Dan Tynan

Kacsmar, Tilbury, Mazmudar, Kerschbaum. Caring about Sharing: User Perceptions of Multiparty Data Sharing. *USENIX Security 2022*

Types of Multiparty Data Sharing



Research Questions

- RQ1: How does the overall acceptability vary across **different types** of multiparty data sharing?
- RQ2: How does acceptability vary in multiparty data sharing for **different user controls** (consent, purpose, retention)?

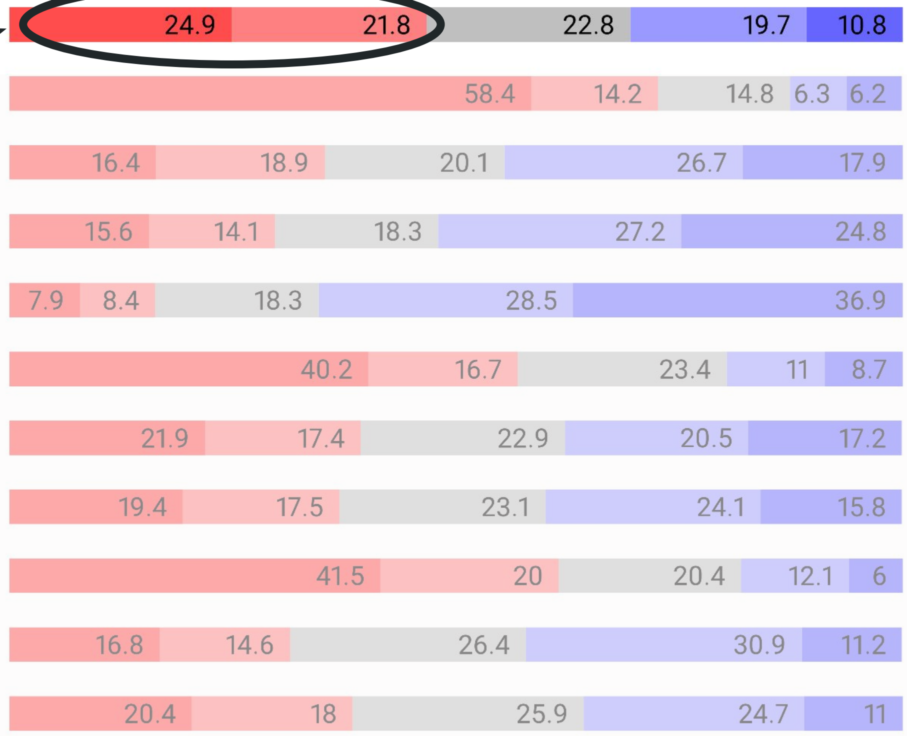
Survey Overview

- 1025 responses through SurveyMonkey in March 2021
- Final participant set is **N = 916**
- Each receives: **1 of 12** scenarios and a series of questions corresponding to user controls
- Use a **five-point semantic differential scale**:

“**Completely Unacceptable**", "Somewhat Unacceptable", “Neutral", “Somewhat Acceptable", “**Completely Acceptable**"

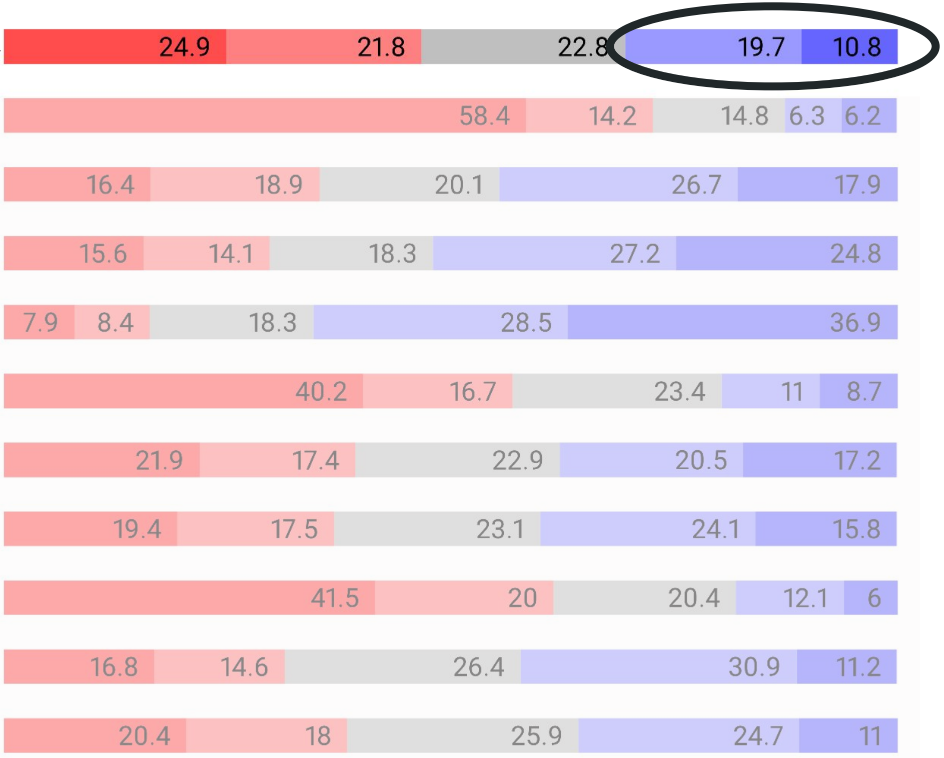
Overall Acceptability Across Scenarios

General Scenario Acceptability?



Overall Acceptability Across Scenarios

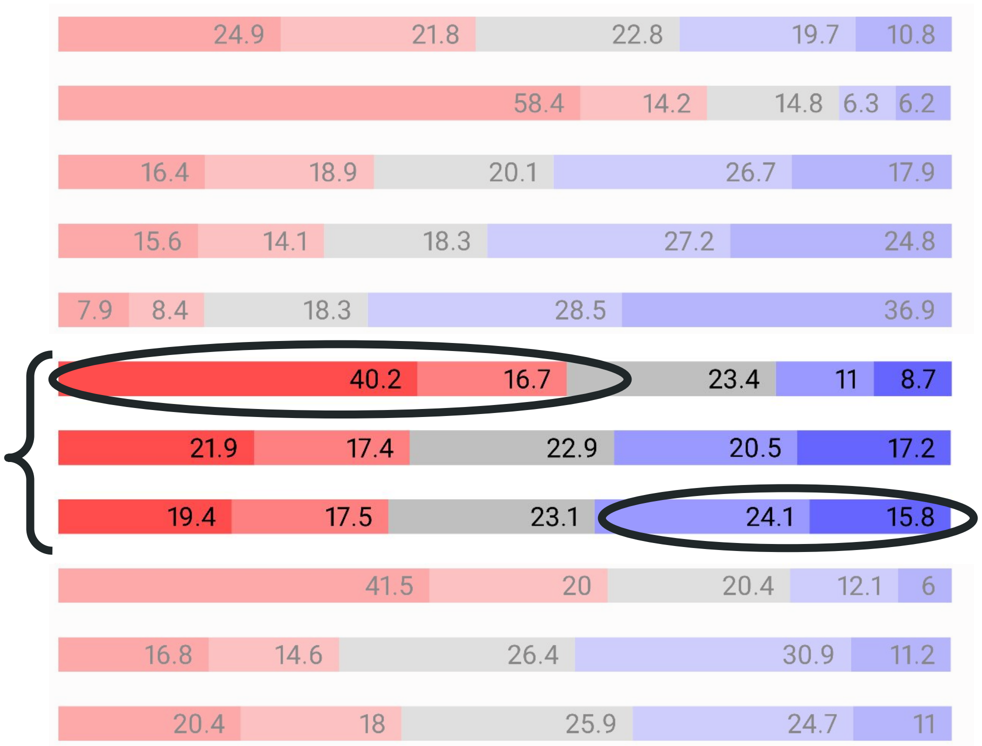
General Scenario Acceptability?



Retention: Acceptability Across All Scenarios

Data Retention?

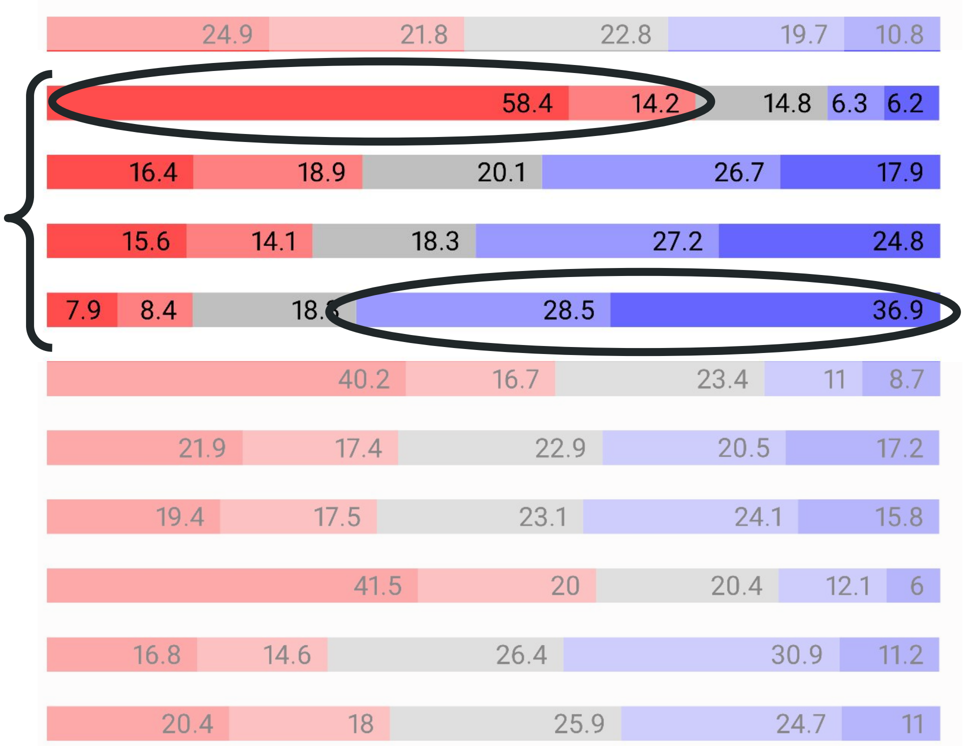
- **Indefinitely**
- **While in use**
- **For set time**



Consent: Acceptability Across All Scenarios

Informed Consent?

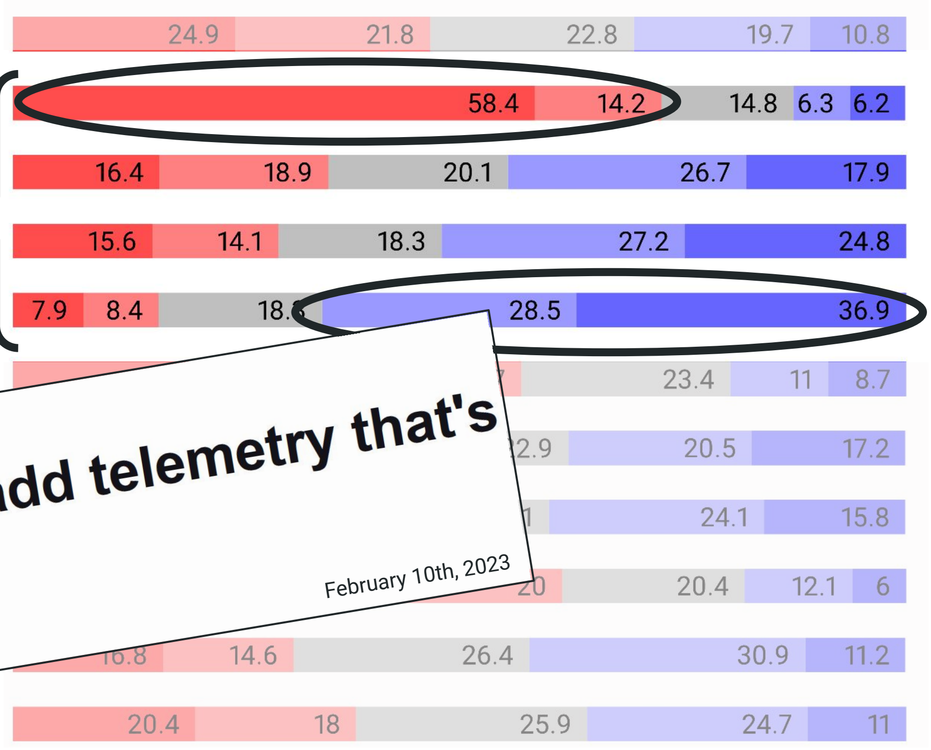
- Concealed 
- Assumed
- Opt-out
- Opt-in 



Consent: Acceptability Across All Scenarios

Informed Consent?

- **Concealed** 
- **Assumed**
- **Opt-out**
- **Opt-in** 



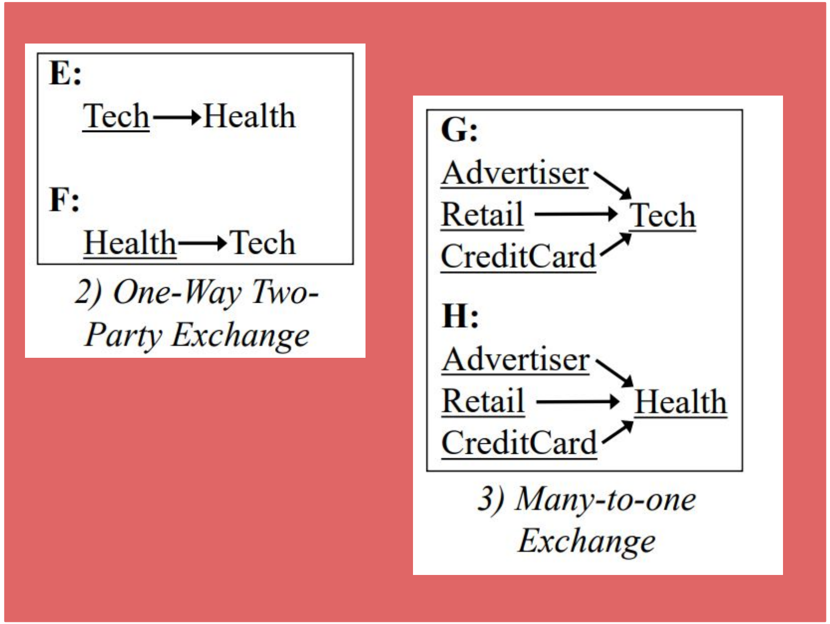
theregister.com

Google's Go may add telemetry that's on by default

February 10th, 2023

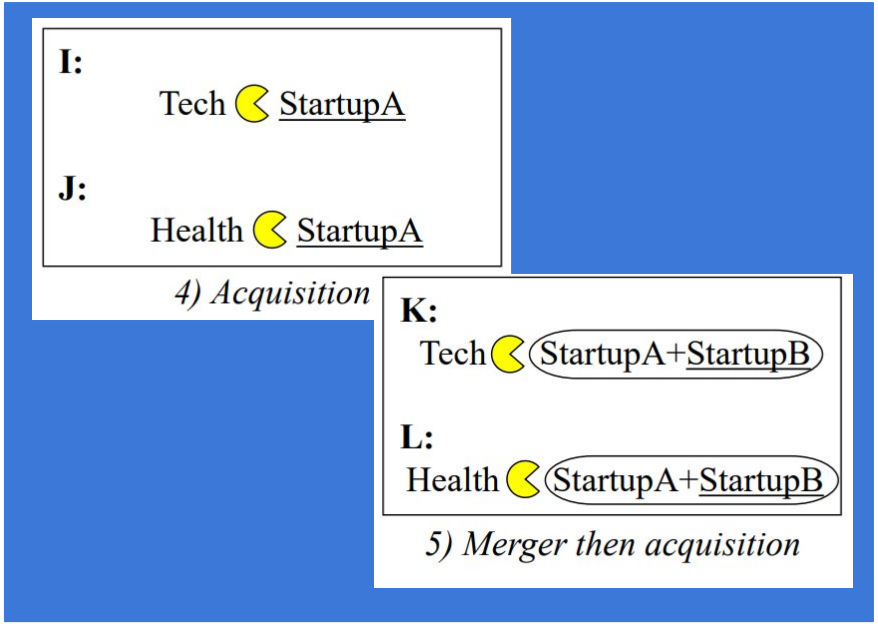
Thomas Claburn

Sharing Type Impact on Overall Acceptability



2) One-Way Two-Party Exchange

3) Many-to-one Exchange



4) Acquisition

5) Merger then acquisition

General acceptability is statistically different between types.

Implications of Sharing Structures

- Disambiguate Third Parties

PetSmart's privacy policy states: "We may share the information we collect with companies that provide support services to us."

- Current systems contains insufficient information to support preferences impacted by sharing type
- Privacy preferences fluctuate with any change to context
- Number of parties, trusted parties, purpose, etc. all influence acceptability, regardless technical privacy