

Beyond Data Privacy for Machine Learning

Bailey Kacsmar

Data, Beyond the Abstraction

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

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



Canada Privacy



0



1,169



11 min read

How Data is Used Continues to Evolve

Microsoft and Providence St. Joseph Health announce strategic alliance to accelerate the future of care delivery - Stories

5-6 minutes

July 8, 2019 | Microsoft News Center

[washingtonpost.com](https://www.washingtonpost.com)

Now for sale: Data on your mental health

Drew Harwell

February 13th, 2023

[cnbc.com](https://www.cnbc.com)

Where Amazon is heading in health after the Amazon Care failure

Eric Rosenbaum



[amazon clinic](https://www.amazon.com/clinic)

Google Health

Privacy matters

When you use Google's products and services, you trust us with your data. It's our responsibility to keep your data private and secure. And at Google Health, we are guided by core privacy and security principles as we build new products and services.

The Use of Health Data

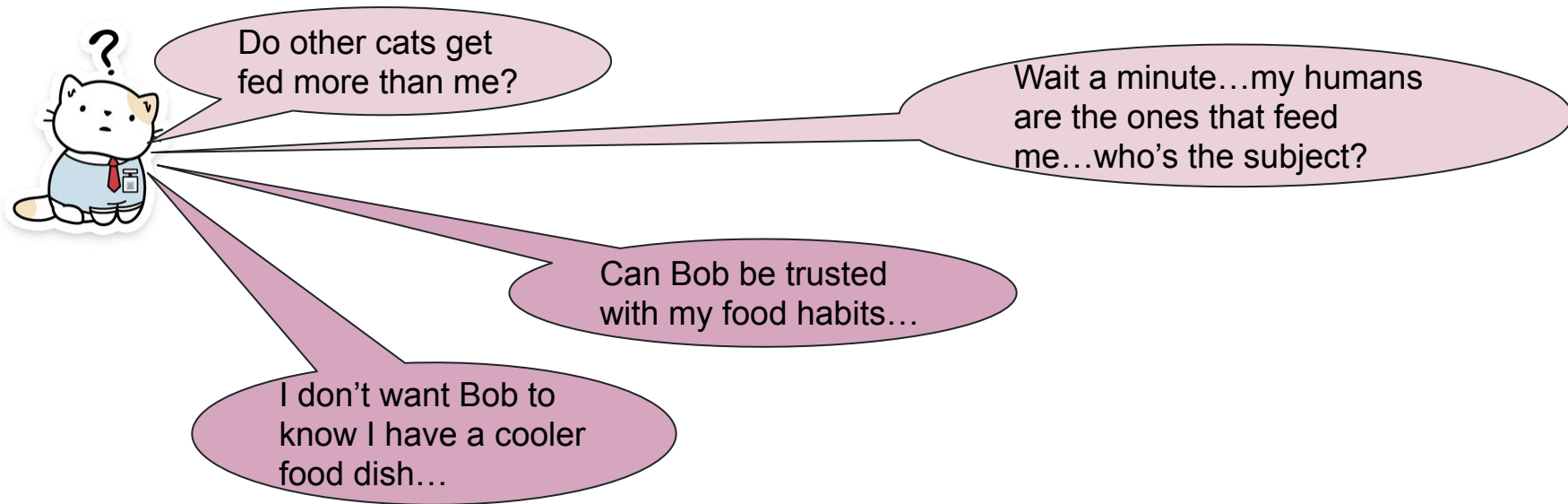


P20986: “It depends. I think it can be beneficial **under certain circumstances**, but I would be hesitant having any healthcare data shared outside my practitioners. However, I recognize how it can improve goods/services, but there **has to be a lot of protection** in place **anytime data is shared**”

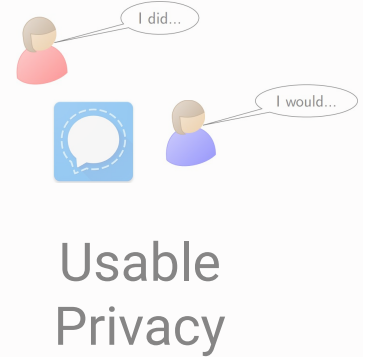
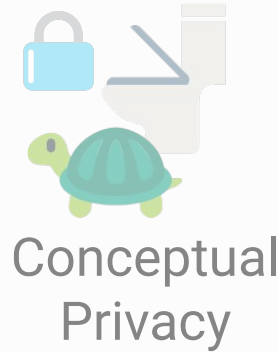
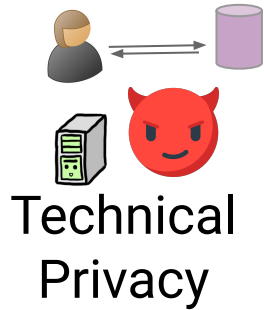
P94865: “**Repugnant**, especially in light of for profit health systems attempting to maximize profitability from patient interactions”



A Technical Privacy Family: Something to Learn

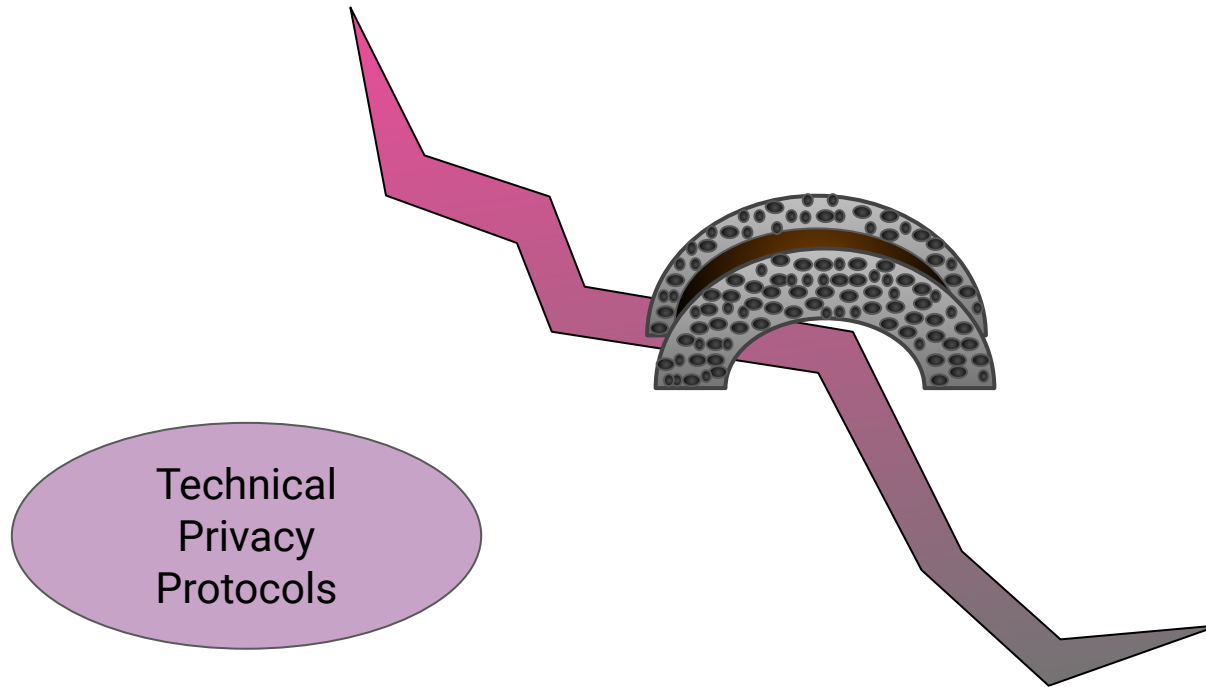


A Wider View of Technical Privacy



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

Technical Solutions for Privacy Problems



Technical Privacy for Machine Learning?

Training Data

Models

Inferences/Outputs

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

Privacy for Machine Learning

Training Data

Models

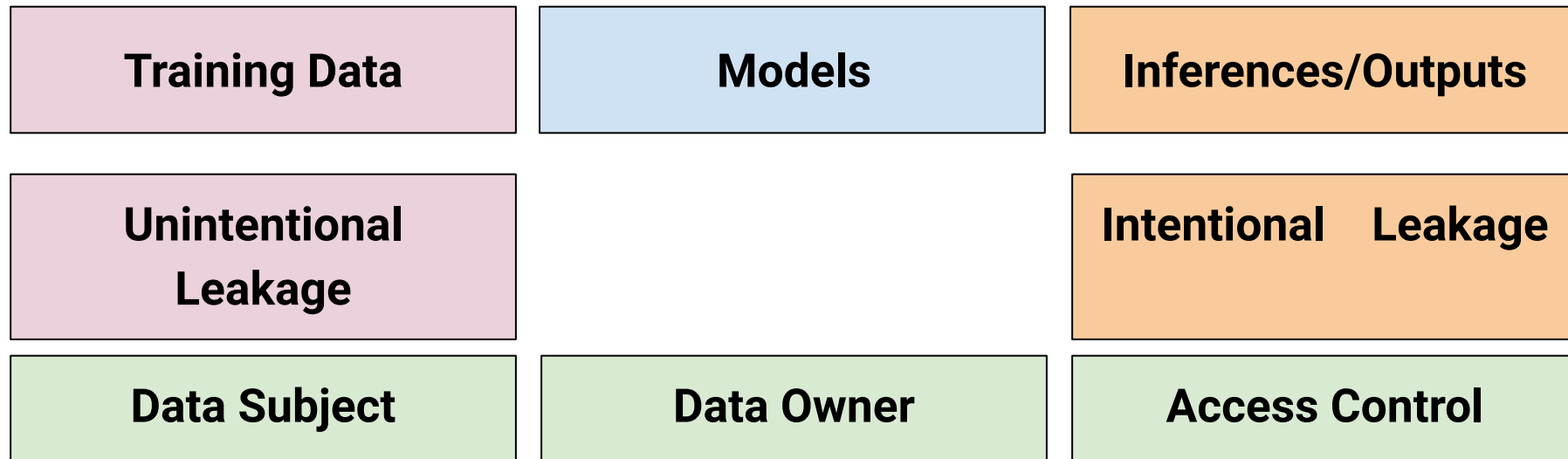
Inferences/Outputs

**Unintentional
Leakage**

Intentional Leakage

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

Privacy for Machine Learning



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

Is this enough?

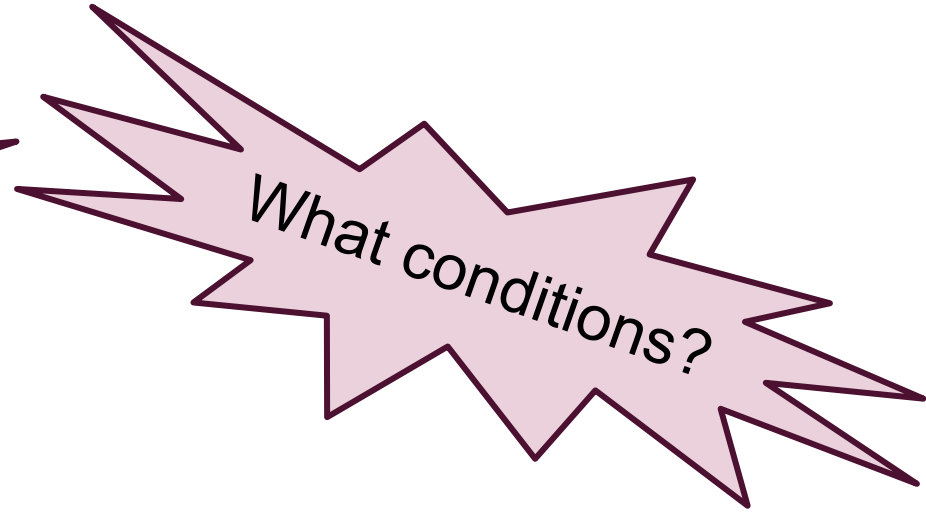
Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Is this enough?



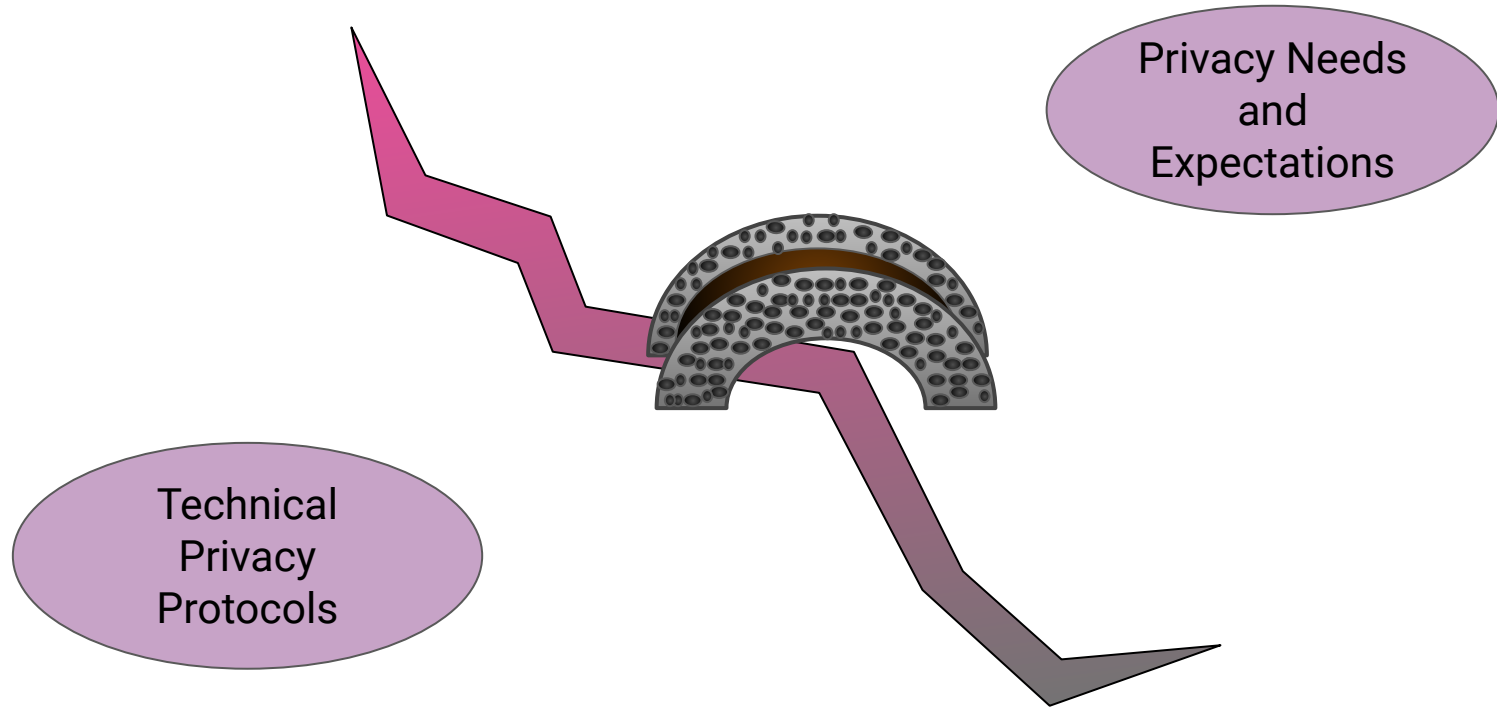
Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Is this enough?

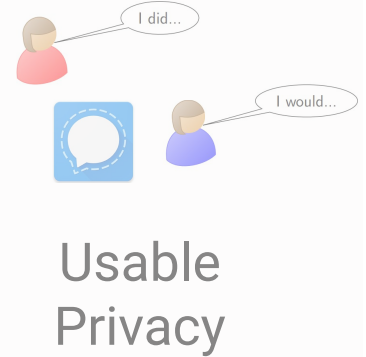
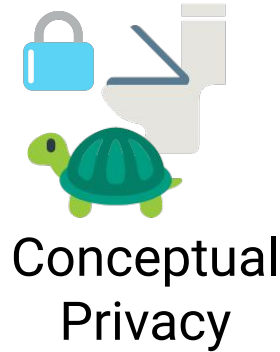
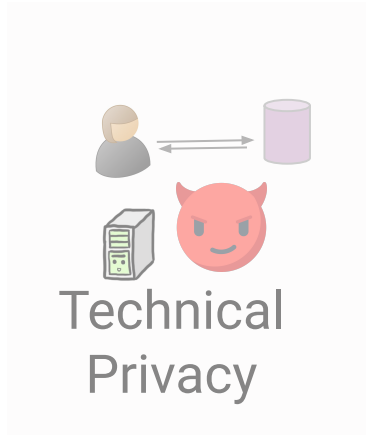


Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Technical Solutions for Privacy Problems



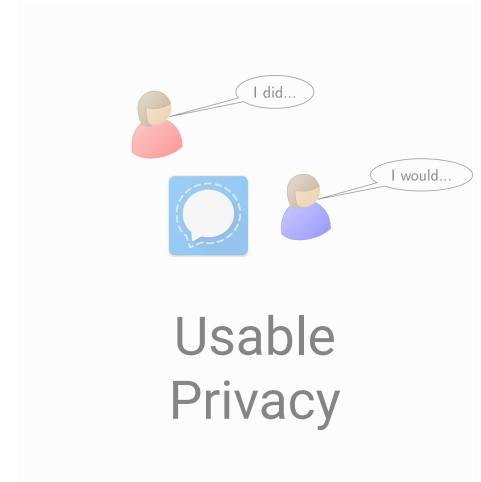
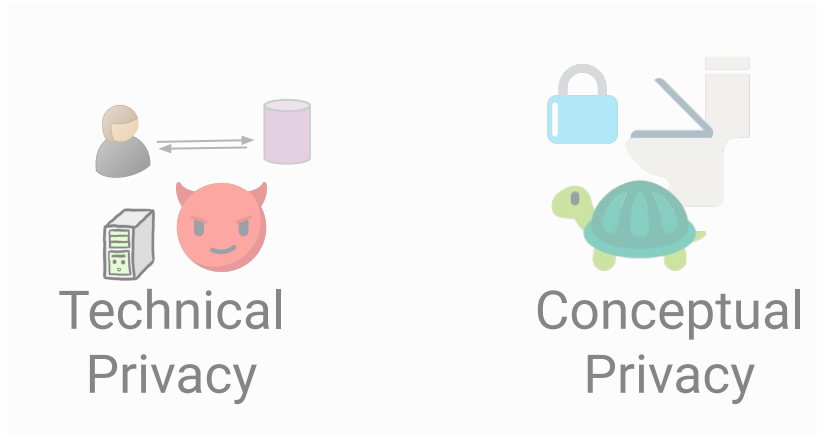
A Wider View of Technical Privacy



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

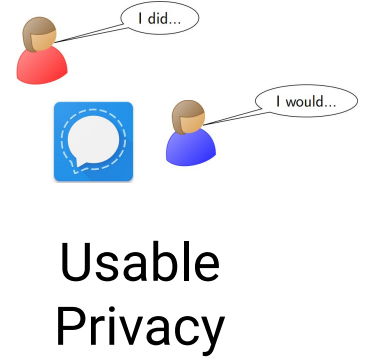
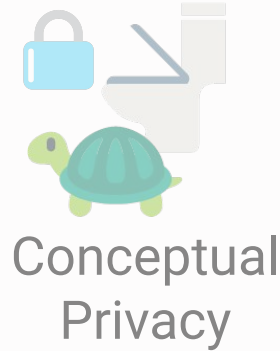
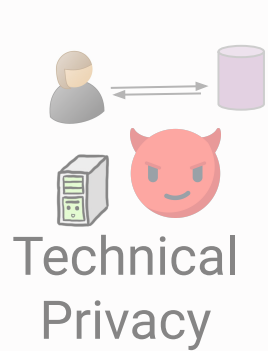
M. Oates, et al. Turtles, locks, and bathrooms: Understanding mental models of privacy through illustration." Proceedings on Privacy Enhancing Technologies 2018.

A Wider View of Technical Privacy



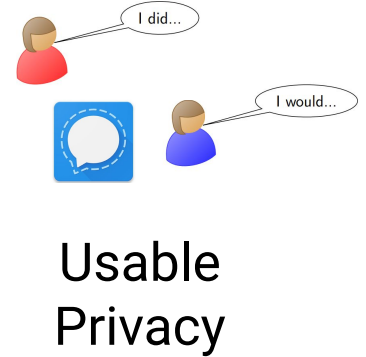
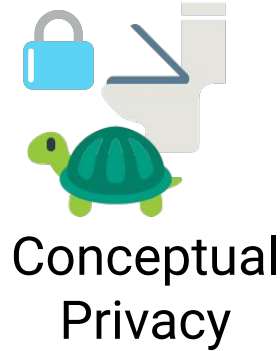
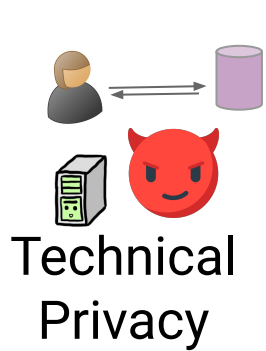
“Trusted-third parties”, “Partners”,

A Wider View of Technical Privacy



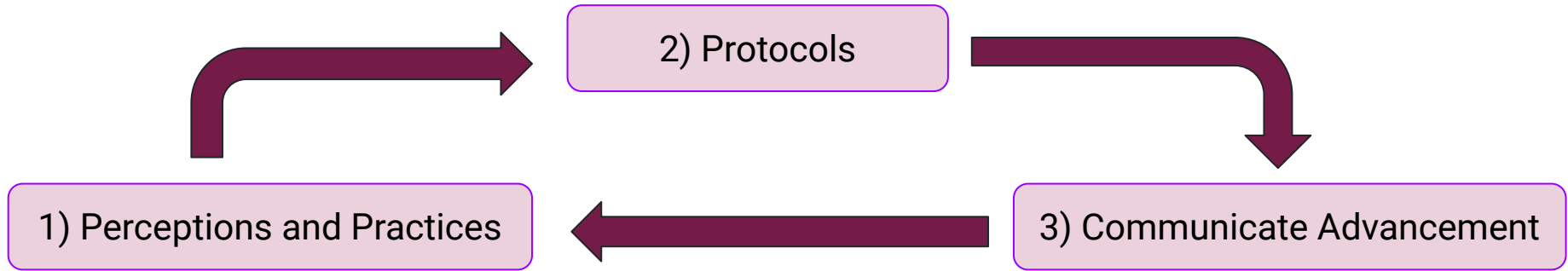
What do users actually do? What do they want to do?

A Wider View of Technical Privacy



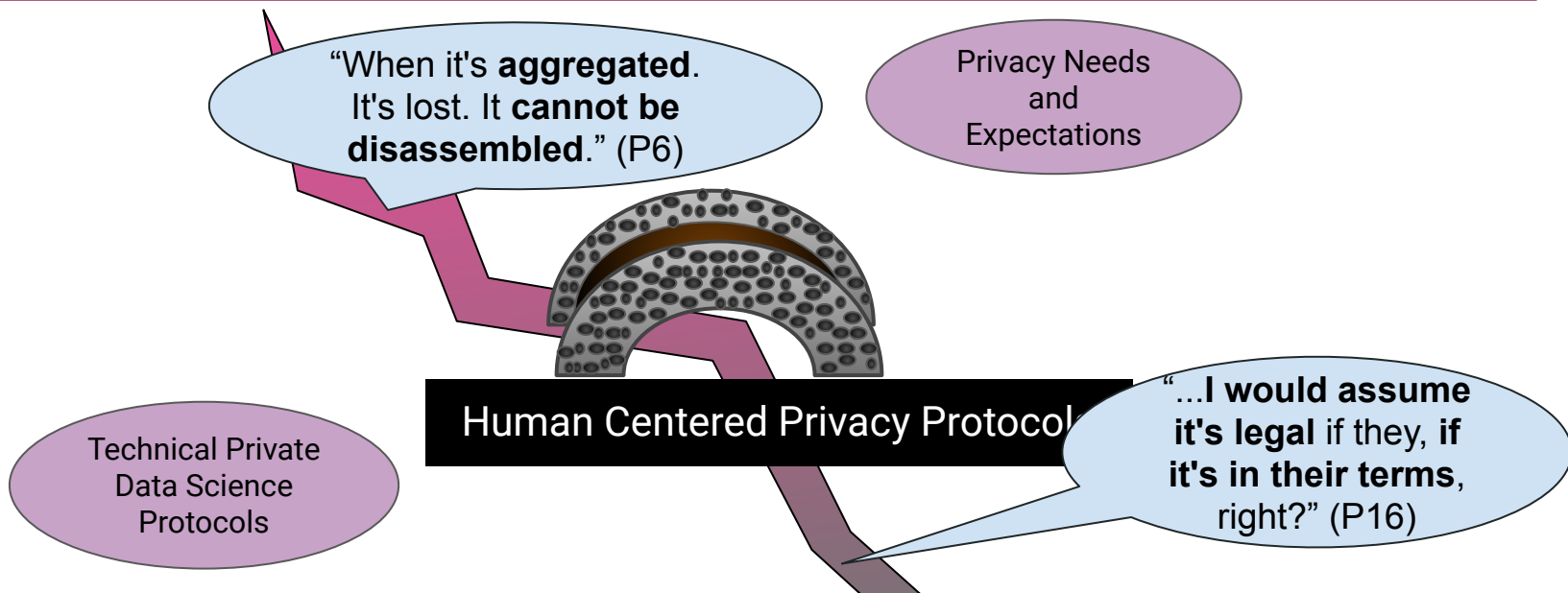
Develop Technical Privacy Solutions Informed by the Breadth of Privacy Notions

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)

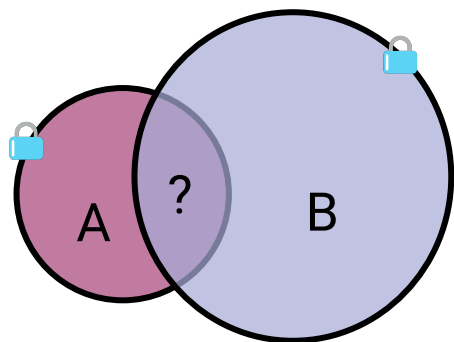
Human Centered Technical Privacy Solutions



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

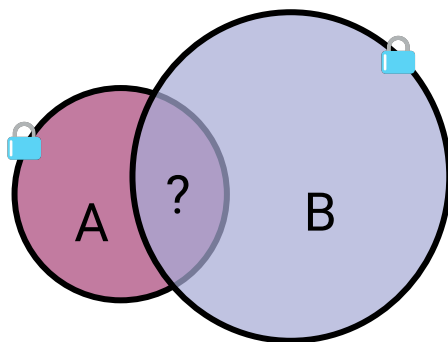
Perceptions of Data Sharing Structures

Structures in Private Computations



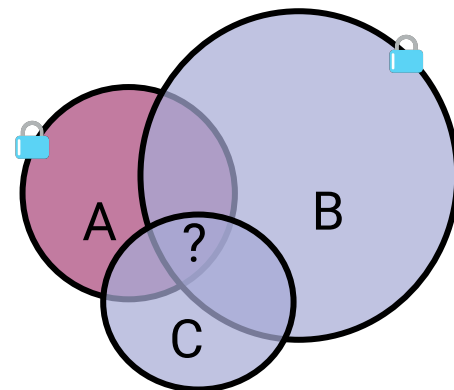
2-Party, One-Way

$A \rightarrow B$



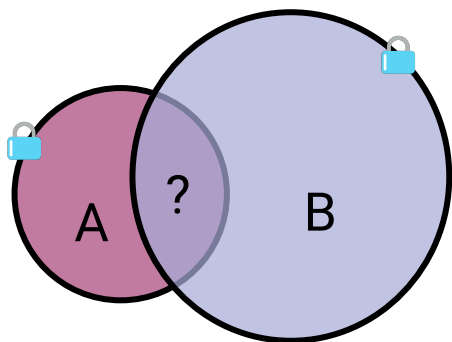
2-Party, Two-Way

$A \leftrightarrow B$



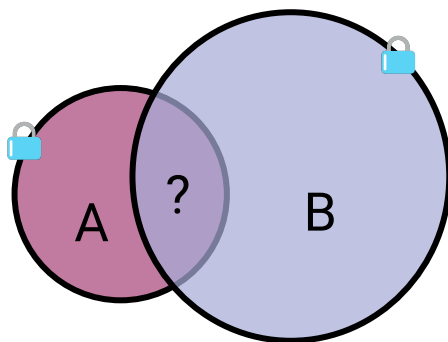
n-Party

Structures in Private Computations



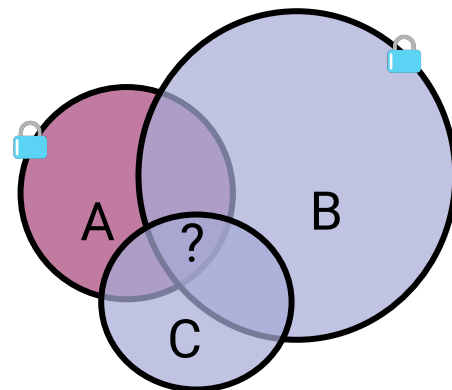
2-Party, One-Way

$A \rightarrow B$



2-Party, Two-Way

$A \leftrightarrow B$



n-Party

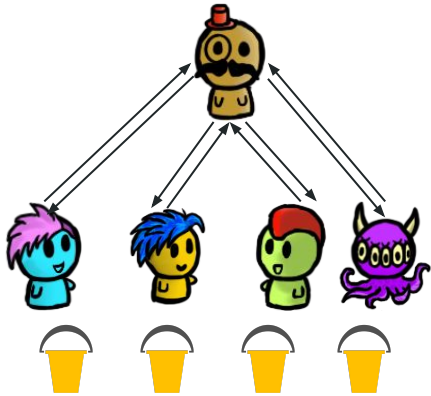
Directionality

Reducing Information

Multi-party

Varying Guarantees

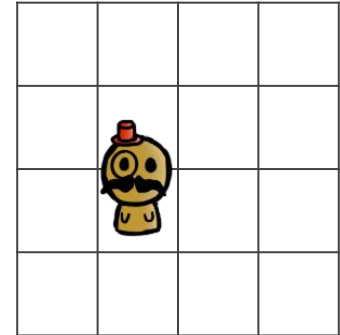
Sample Sharing Structures in Machine Learning



Federated Learning



Non-Federated Learning



Reinforcement Learning

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

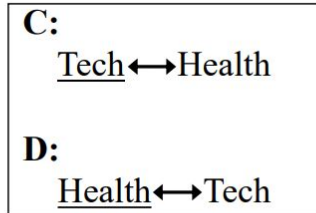
What happens to your data when a company dies? - The Parallax

Dan Tynan

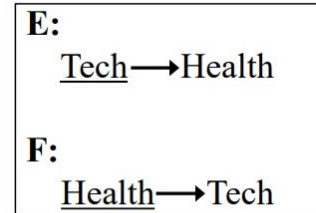
Types of Multiparty Data Sharing



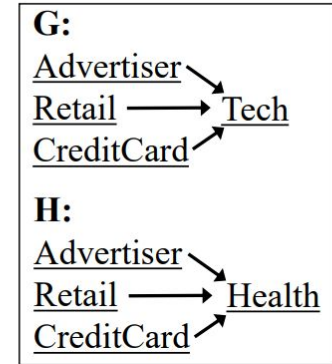
V) Validation



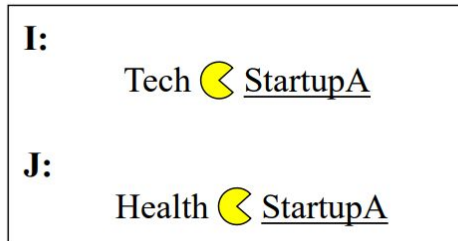
1) Two-Way Two-Party Exchange



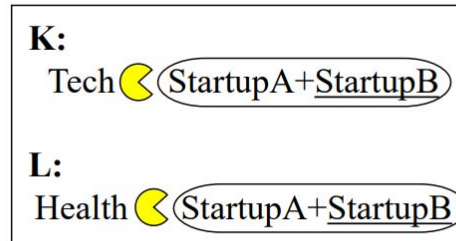
2) One-Way Two-Party Exchange



3) Many-to-one Exchange



4) Acquisition



5) Merger then acquisition

$X \rightarrow Y$: X provides data to Y

$X \leftrightarrow Y$: X and Y provide data to each other

$X \text{ ☾ } Y$: X acquires Y

$(X+Y)$: X merges with Y

X: scenario indicated you are a user of X

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

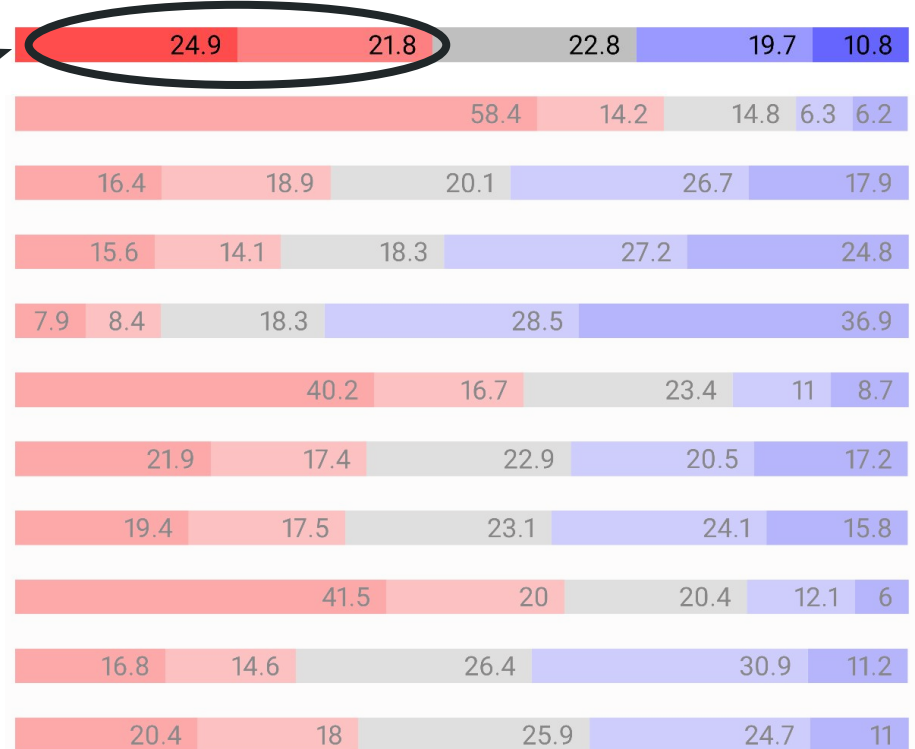


- 1025 responses through SurveyMonkey in March 2021
- Final participant set is **N = 916**
- Each receives: **1 of 12** scenarios and a series of questions corresponding to user controls
- Use a **five-point semantic differential scale**:

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

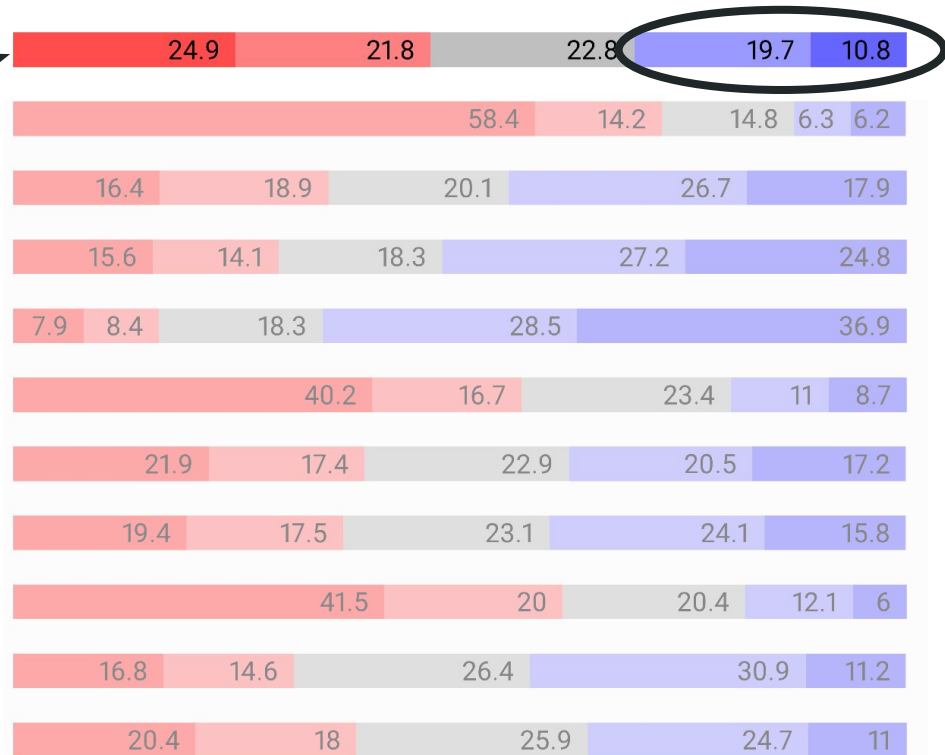
Overall Acceptability Across Scenarios

**General Scenario
Acceptability?**



Overall Acceptability Across Scenarios

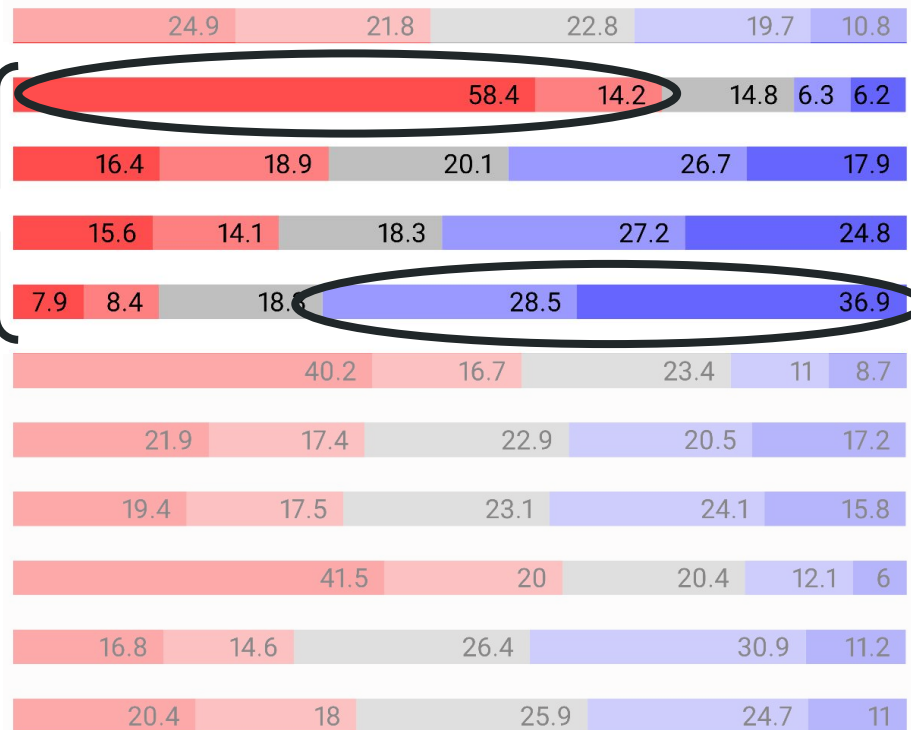
**General Scenario
Acceptability?**



Consent: Acceptability Across All Scenarios



Informed Consent?

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

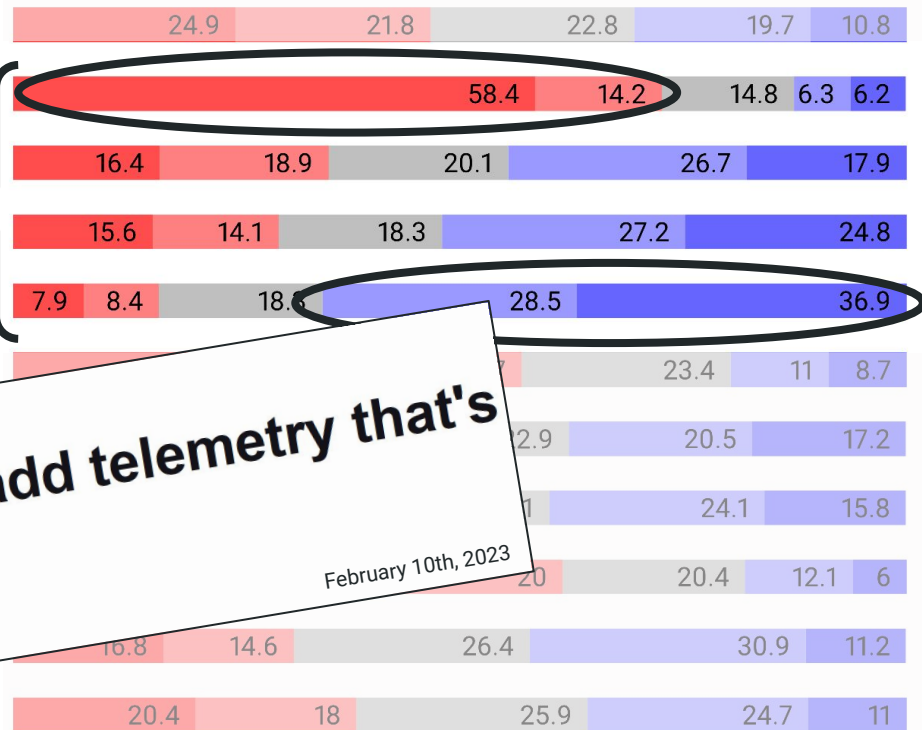


Consent: Acceptability Across All Scenarios

Informed Consent?

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

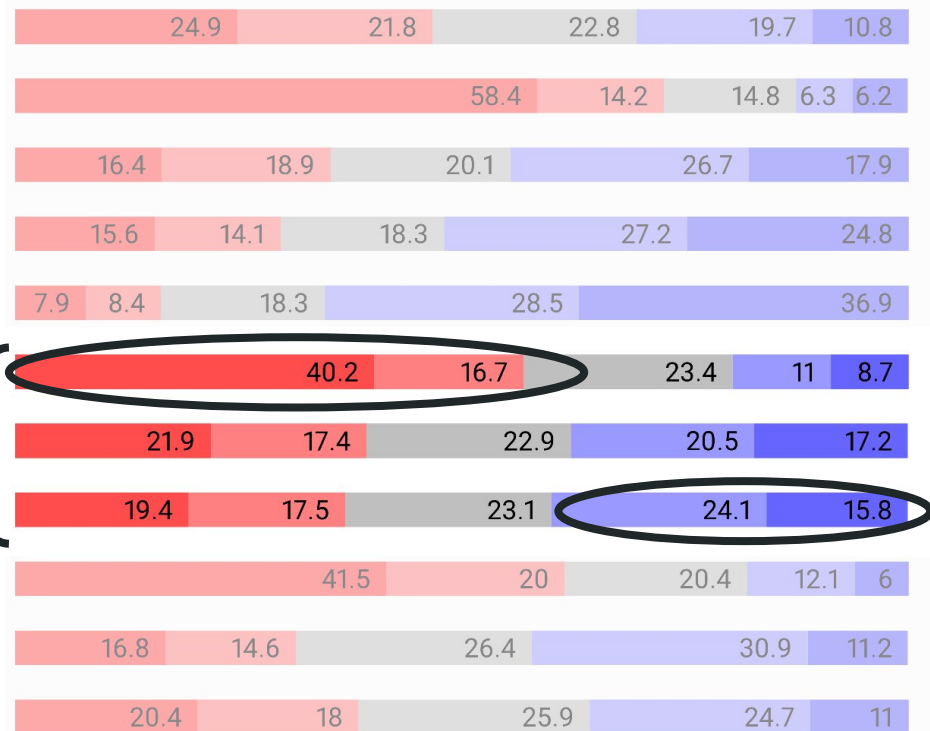
[theregister.com](https://www.theregister.com)
Google's Go may add telemetry that's on by default
February 10th, 2023
Thomas Claburn



Retention: Acceptability Across All Scenarios

Data Retention?

- Indefinitely
- While in use
- For set time



Sharing Type Impact on Overall Acceptability

E:
Tech → Health

F:
Health → Tech

2) One-Way Two-Party Exchange

G:
Advertiser → Tech
Retail → Tech
CreditCard → Tech

H:
Advertiser → Health
Retail → Health
CreditCard → Health

3) Many-to-one Exchange

I:
Tech ☾ StartupA

J:
Health ☾ StartupA

4) Acquisition

K:
Tech ☾ (StartupA+StartupB)

L:
Health ☾ (StartupA+StartupB)

5) Merger then acquisition

General acceptability is statistically different between types.

Implications of Sharing Structures

- Disambiguate Third Parties

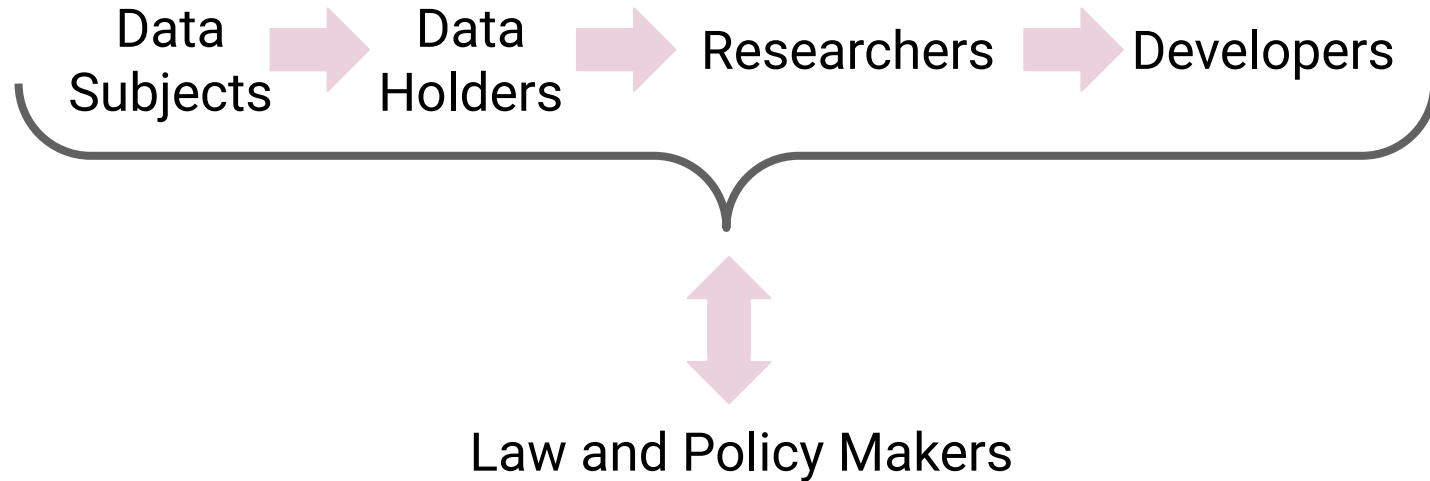
PetSmart's [privacy_policy](#) states: "We may share the information we collect with companies that provide support services to us."

- 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



Communicating Privacy

Privacy Perceptions and Expectations



Interview Study

- 22 participants
- Average 60 minutes (longest 90, shortest 40)
- Recruited via prolific
- Interviews were done online
- Participants were located across the United States

Research Questions

- **RQ1:** What do data subjects understand about private computation, and how can specific examples facilitate their understanding of the concept?
- **RQ2:** How is a data subjects' willingness to share their data impacted when informed of private computation's properties (protections and guarantees)?
- **RQ3:** How do data subjects perceive private computation's risks (e.g., inference attacks and beyond)?



Participant Comprehension and Expectations

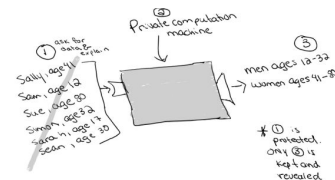


First Attempt



Second Attempt

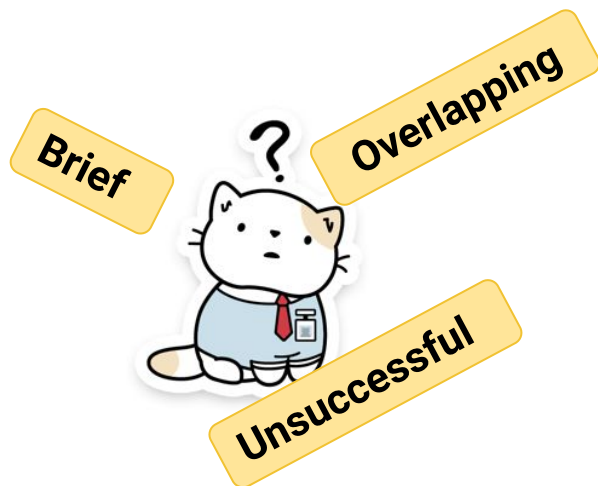
Secure computation is a way that a company analyzes your data. The final analysis will be made public [at access location]. However, your specific data is protected and cannot be traced back to you nor can your specific data points be traced back to you. The analysis will be specifically [example], and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

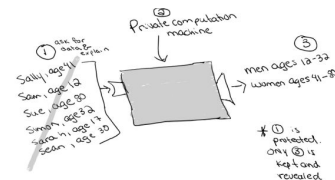


First Attempt



Second Attempt

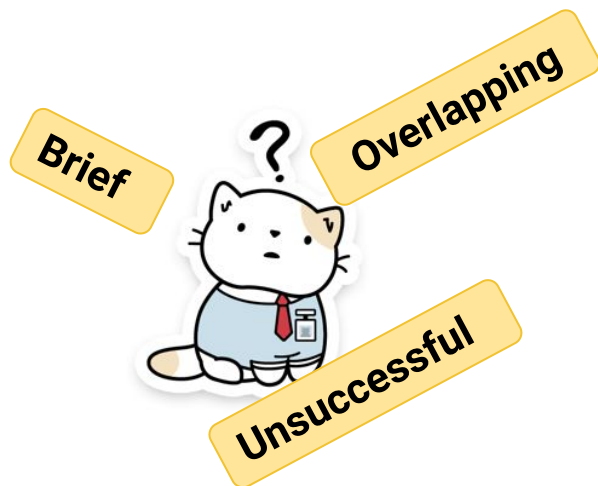
Secure computation is a way that a company analyzes your data. The final analysis will be made public [at access location]. However, your specific data is protected and cannot be traced back to you nor can your specific data points be traced back to you. The analysis will be specifically [example], and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

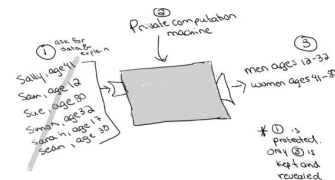


First Attempt



Second Attempt

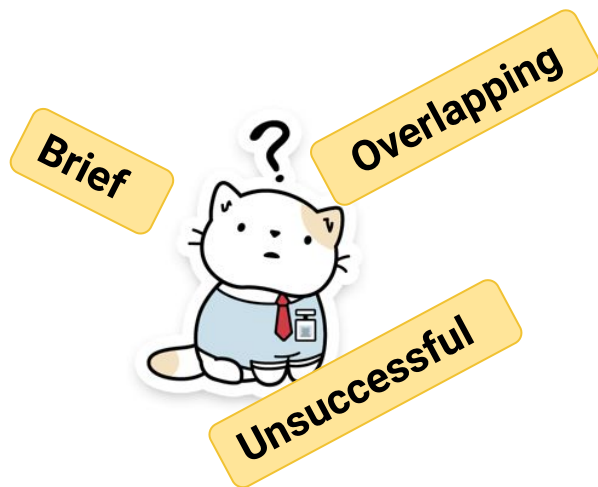
Secure computation is a way that a company analyzes your data. The final analysis will be made public [at access location]. However, your specific data is protected and cannot be traced back to you nor can your specific data points be traced back to you. The analysis will be specifically [example], and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

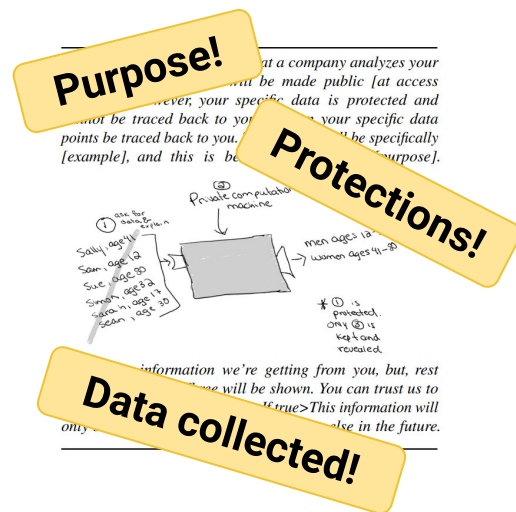
Participant Comprehension and Expectations



First Attempt

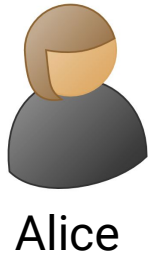


Second Attempt

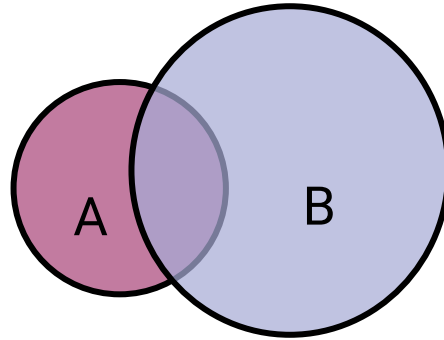


Final Consensus

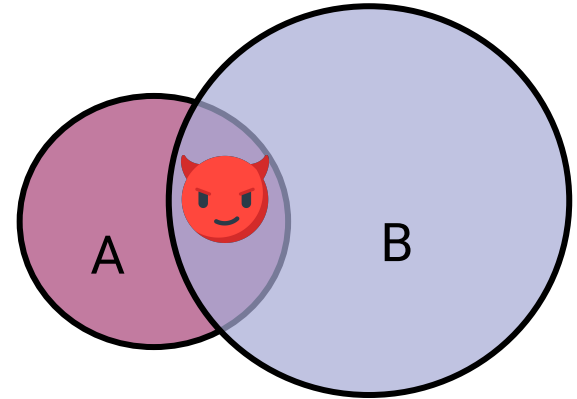
Awareness of Unique Threat Models



Joins Social App



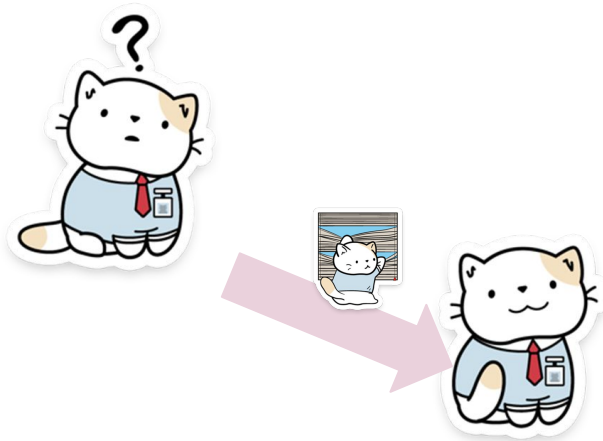
Contact Discovery



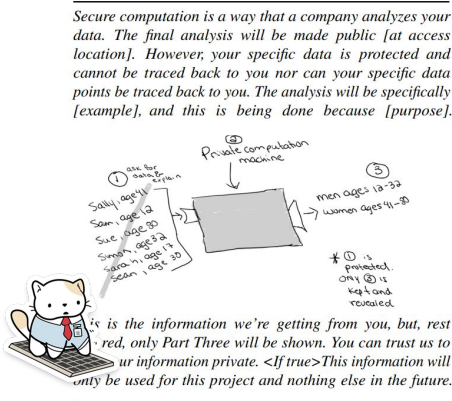
Real Identity Connected

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

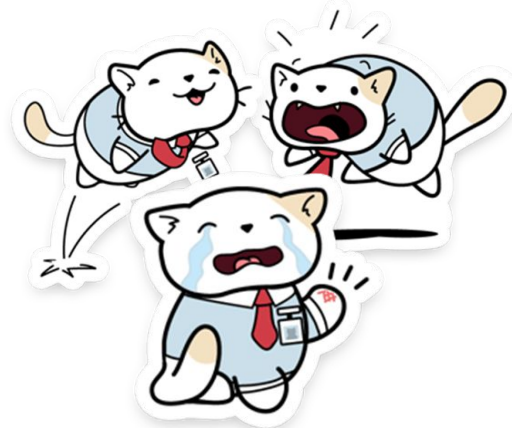
Communication Results in a Snapshot



Built Participant
Comprehension



Key Information to
Provide to Data Subjects



Awareness of Unique
Threat Models

People can reason about private computation; let them

Privacy for Machine Learning

- Technical solutions are a powerful tool for protecting data
- For protections to correspond to personal privacy, we need to know the expectations
- **Protections provided by protocols** and constructions **do not encompass the full range of risks** experienced **by individuals** in society

But, a wider view of technical privacy allows us to provide **better protections** against risks that are in range.

Privacy for Machine Learning

- Technical solutions are a powerful tool for protecting data
- For protections to correspond to personal privacy, we need to know the expectations
- **Protections provided by protocols and constructions do not encompass the full range of risks experienced by individuals** in society

But, **a wider view of technical privacy** allows us to provide **better protections** against risks that are in range.

Thanks!

Bonus!

Turtles, locks, and bathrooms



PEARL OYSTERS HAVE SOMETHING VALUABLE
TO PROTECT - THE PEARL.
THEY CAN DO SO BY SIMPLY CLOSING THE LID.
IF ONLY SAFEGUARDING THE DATA IN MY
LAPTOP WERE THAT SIMPLE!

Fig. 62. “Pearl oysters have something valuable to protect - the pearl. They can do so by simply ‘closing the lid.’ If only safeguarding the data in my laptop were that simple!” By Sharon, age 25.



Fig. 33. “Privacy means that the thoughts in my brain are locked away. What I know does not have to go into the world, which I put an X over.” By Thomas, age 19



Fig. 24. “No one come in when I am in the bathroom!” By Sydney, age 7

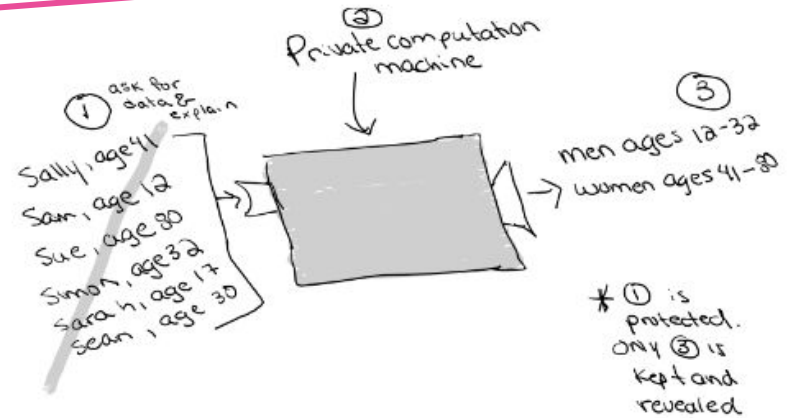


Fig. 23. “This is me enjoying my privacy. This is the only time during the day, were I am truly alone and nothing bothers me. No man no children no dogs.” By Cindy, age 54

Why are you doing this?
What is your motivation?

“Want to determine whether [...] **their ads are effective? Well, you're still in business right?** See, that for me, that's enough.” (P16)

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].

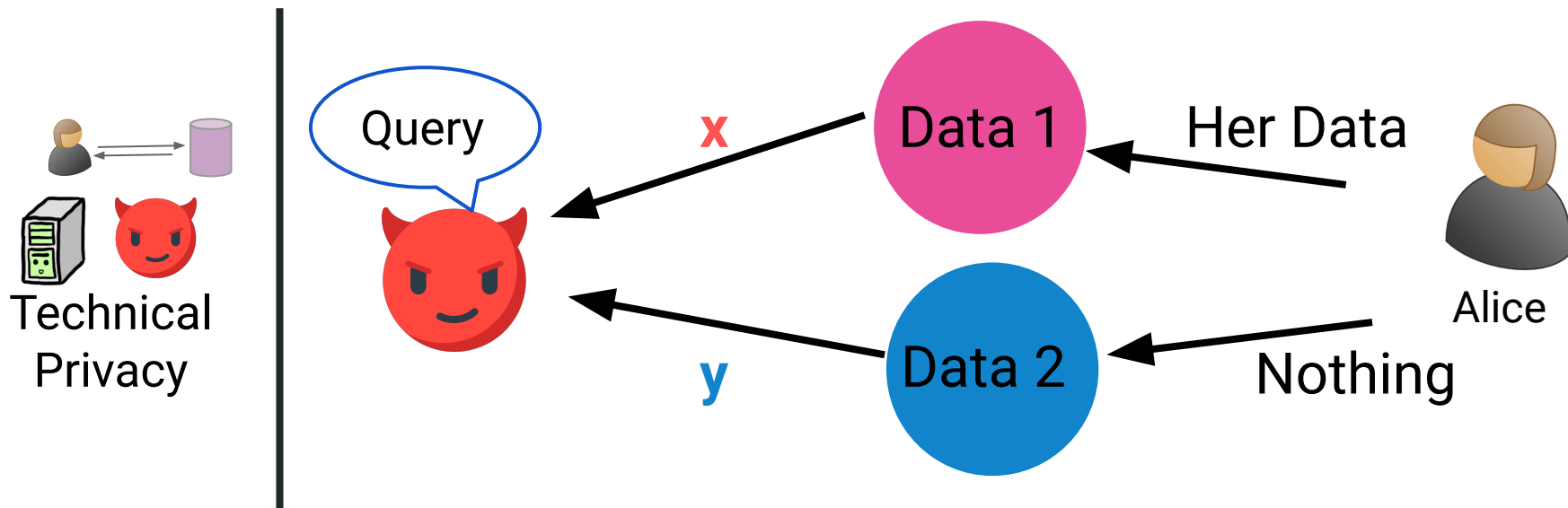


This is the information we're getting from you, but, rest assured, only Part Three will be shown. You can trust us to keep your information private. <If true>This information will only be used for this project and nothing else in the future.

Scenario C

TechForYou is a large internet company that offers a search engine, email accounts and smartphone platforms to users. GoodHealth runs a chain of hospitals across the country and stores health data for millions of patients during its day-to-day operations. TechForYou and GoodHealth **will share the customer data they hold with one another**. You are a customer of TechForYou. How acceptable is this scenario?

Technical Privacy: Differential Privacy Intuition

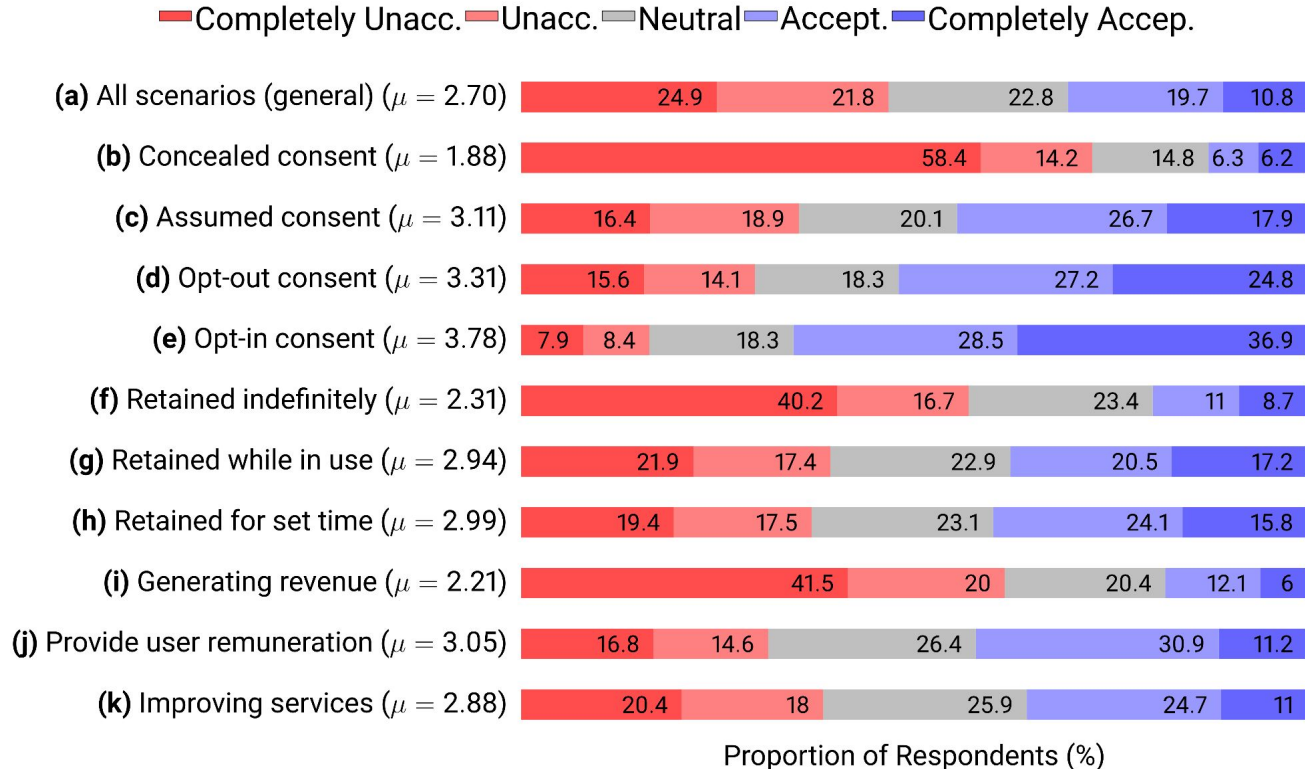


Define, **what** is being protected, from **who**, and under what **conditions** this protection will hold.

Average Acceptability Within Sharing Type

	Type 1	Type 2	Type 3	Type 4	Type 5	Average Acceptability Score
(a) All scenarios	2.63	2.51	2.34	2.96	2.93	
(b) Concealed consent	1.96	1.71	1.77	1.84	2.00	
(c) Assumed consent	3.00	2.91	2.99	3.34	3.34	
(d) Opt-out consent	3.20	3.15	3.19	3.53	3.49	
(e) Opt-in consent	3.71	3.63	3.69	4.00	3.96	
(f) Retained indefinitely	2.34	2.27	2.10	2.42	2.51	
(g) Retained while in use	3.04	2.79	2.87	2.95	2.87	
(h) Retained set time	3.12	2.72	2.81	3.17	3.07	
(i) Generating revenue	2.14	2.04	2.27	2.23	2.36	
(j) Provide user remuneration	3.02	2.85	3.11	3.21	3.13	
(k) Improving services	2.81	2.70	2.78	2.91	3.11	

Acceptability Distribution Across All Scenarios



Non-Transferable, Free, and Transparent Consent

P09262: “...**specific consent** is received from the customer to **where/what** the information is shared to, as well as **why**”

P41281: “Information collected, with the users permission, **should never be shared** with another company **or assumed to be the property** of said company if they merge with another company...”

P66884: “It’s inappropriate unless the user consents **explicitly** and should **never be a requirement for use**”

Polarizing



P58310: I think companies after having **acquired data as an asset** has one intention and it's making money through **exploitation**"

P20322: "I'm not happy about it because **if you** do **agree you can't choose** who it will be shared with.
If you don't agree, you can't use the service"



P14505: "I think that it is **acceptable** because **they need to use this data** for advertising opportunities"