Towards Designing Technical Solutions for Privacy Laws

Kobbi Nissim
Georgetown University

The Data Co-Ops project
https://datacoopslab.org

[NASEM, May 2023]
[partly based on joint w/ M. Altman & A. Cohen]

A sense of urgency

- An extremely large (and growing) number of decisions of legal consequence are made in computer systems …

- Credit score
- Assisting legal decision making
- Assisting hiring decisions
- Academic performance evaluation
- Online advertising and personalized content delivery

- … even if only a small fraction required human review, they would quickly overwhelm judiciary or administrative systems

A sense of urgency

- An extremely large (and growing) number of decisions of legal consequence are made in computer systems …

- Credit score
- Assisting legal decision making
- Assisting hiring decisions
- Academic performance evaluation
- Online advertising and personalized content delivery

- … even if only a small fraction required human review, they would quickly overwhelm judiciary or administrative systems

- Systems design must ensure that such decisions would (almost) always be done right

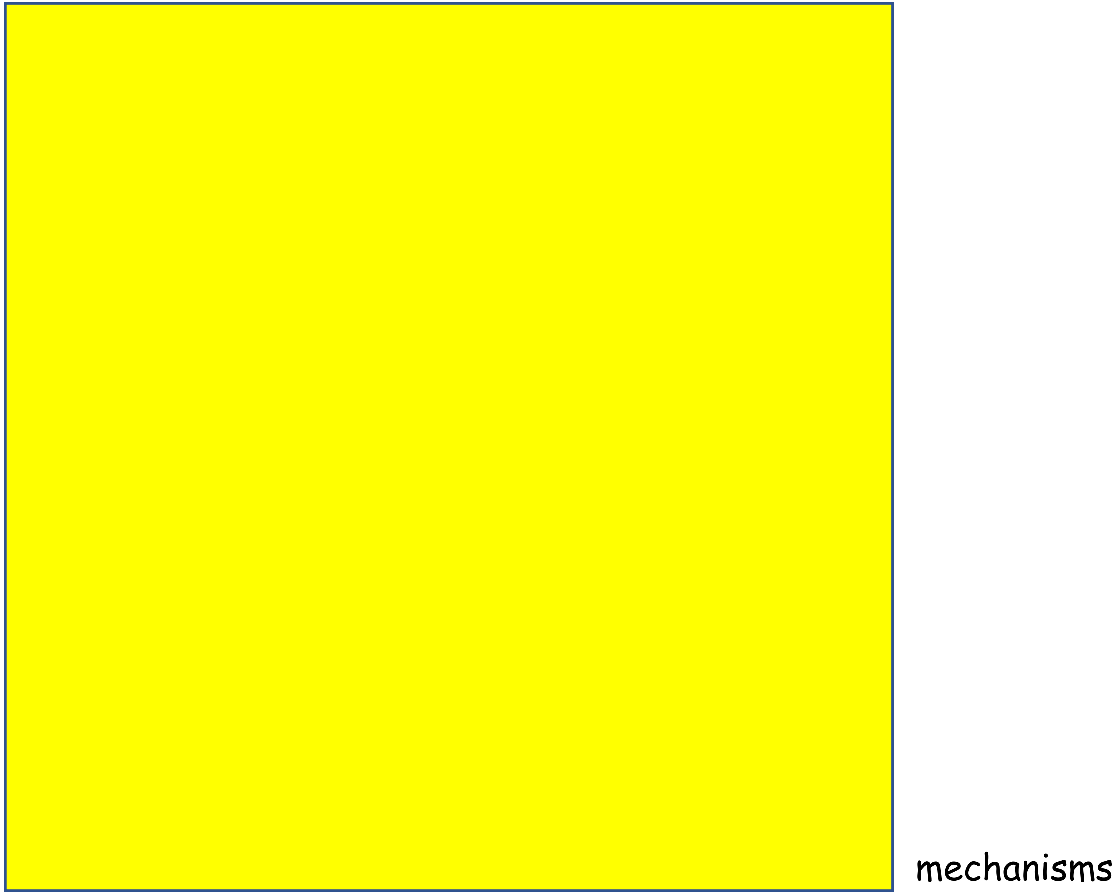
How to design technical systems that meet privacy laws?

How to design technical systems that meet privacy laws?

Can we at least map legal data protection requirements to technical specifications?

A CS interpretation of a legal privacy standard

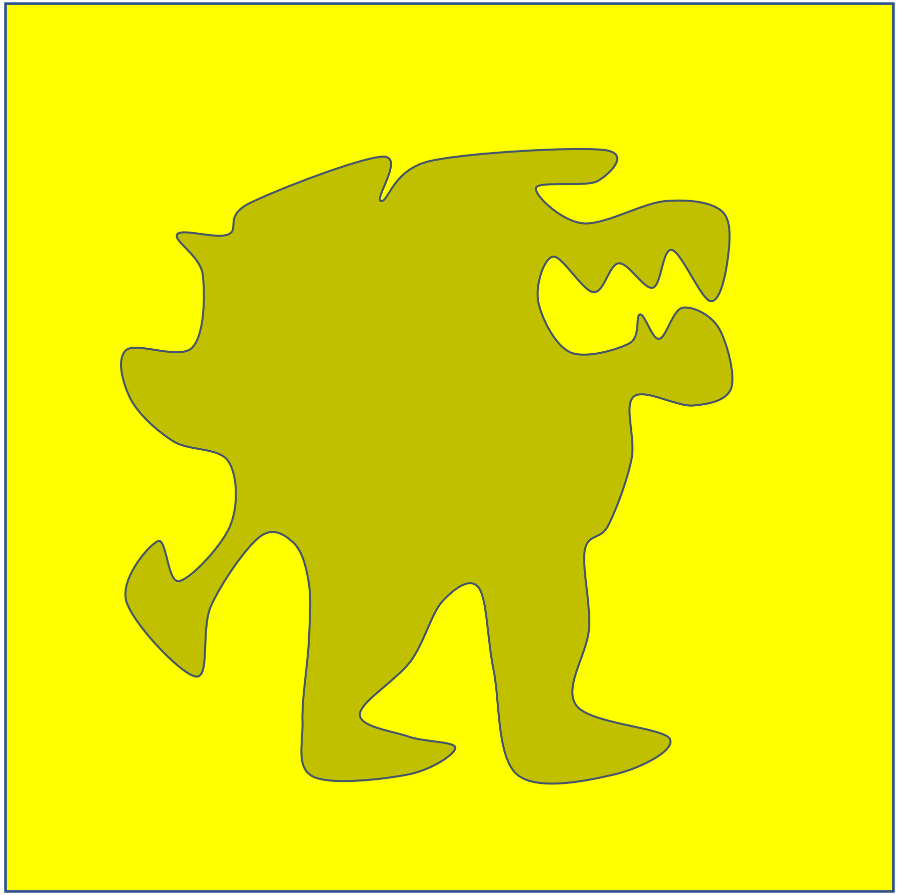
A CS interpretation of a legal privacy standard



A CS interpretation of a legal privacy standard

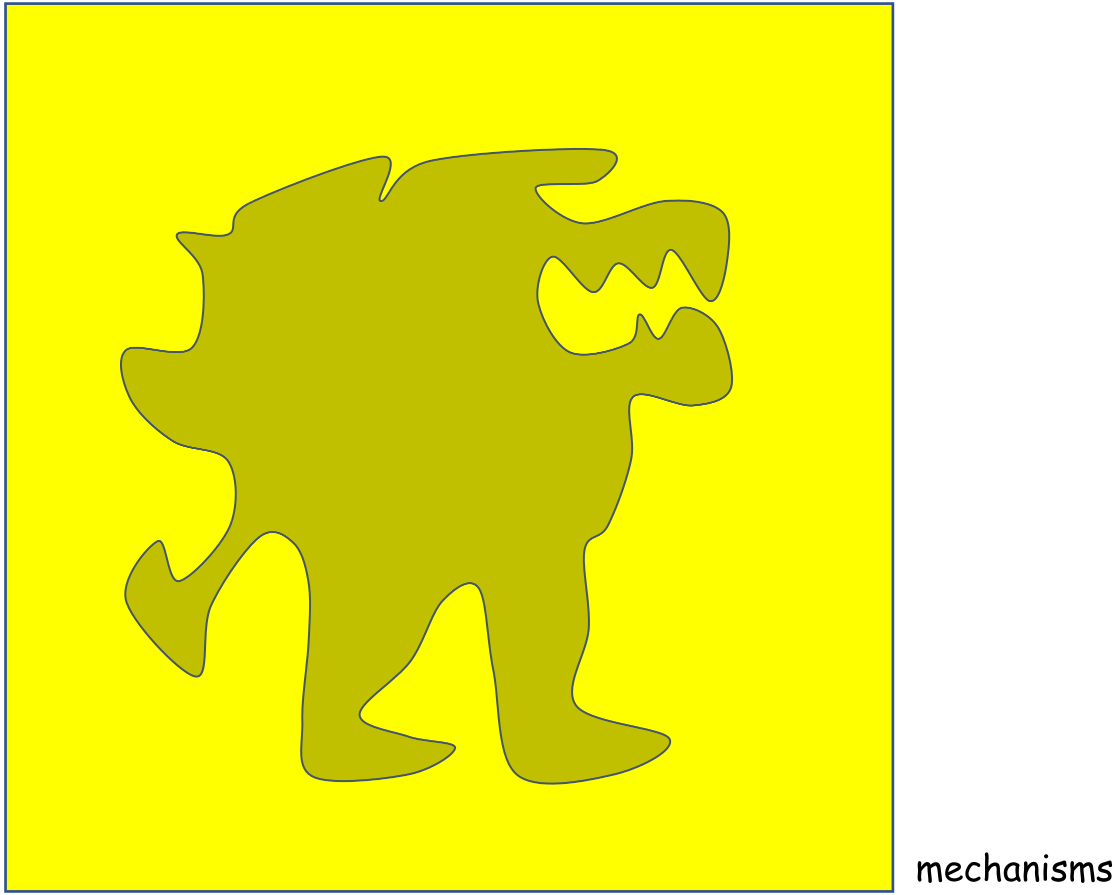
mechanisms

A CS interpretation of a legal privacy standard



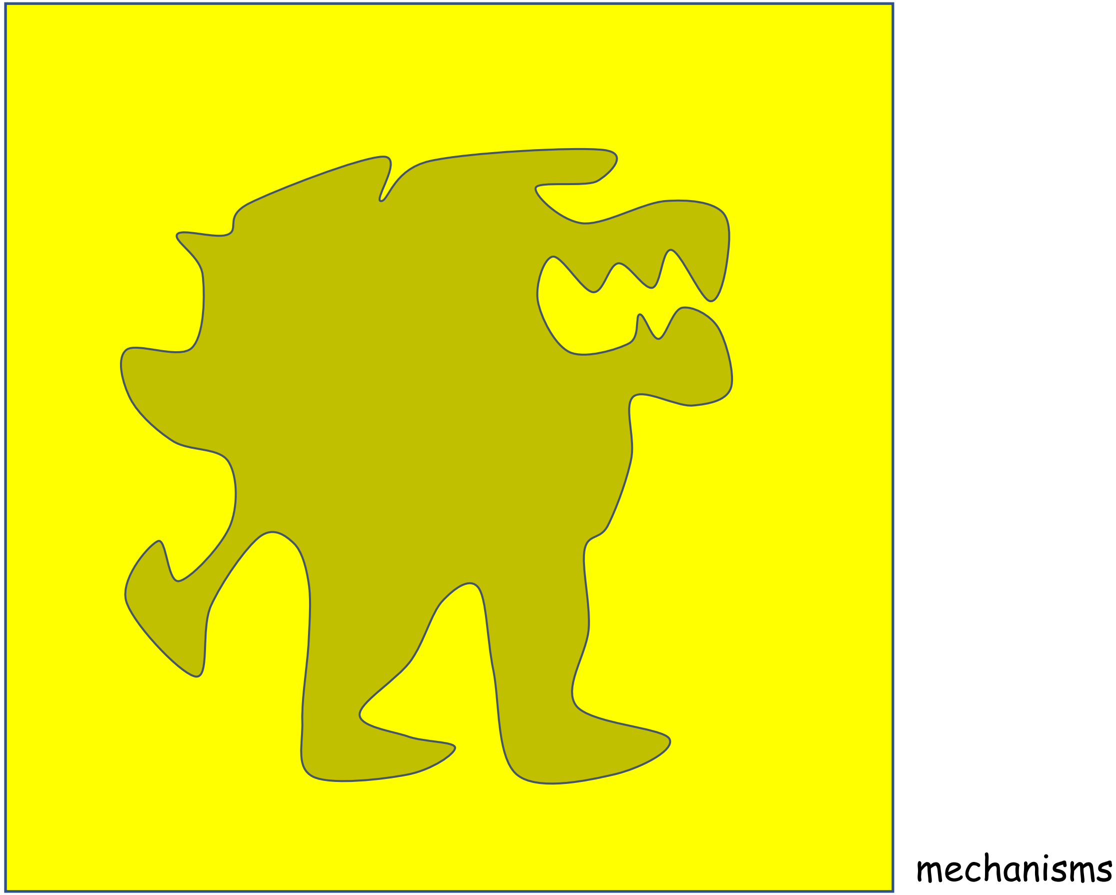
A CS interpretation of a legal privacy standard

- How can I design systems without having a clear definition of what I am supposed to do?



A CS interpretation of a legal privacy standard

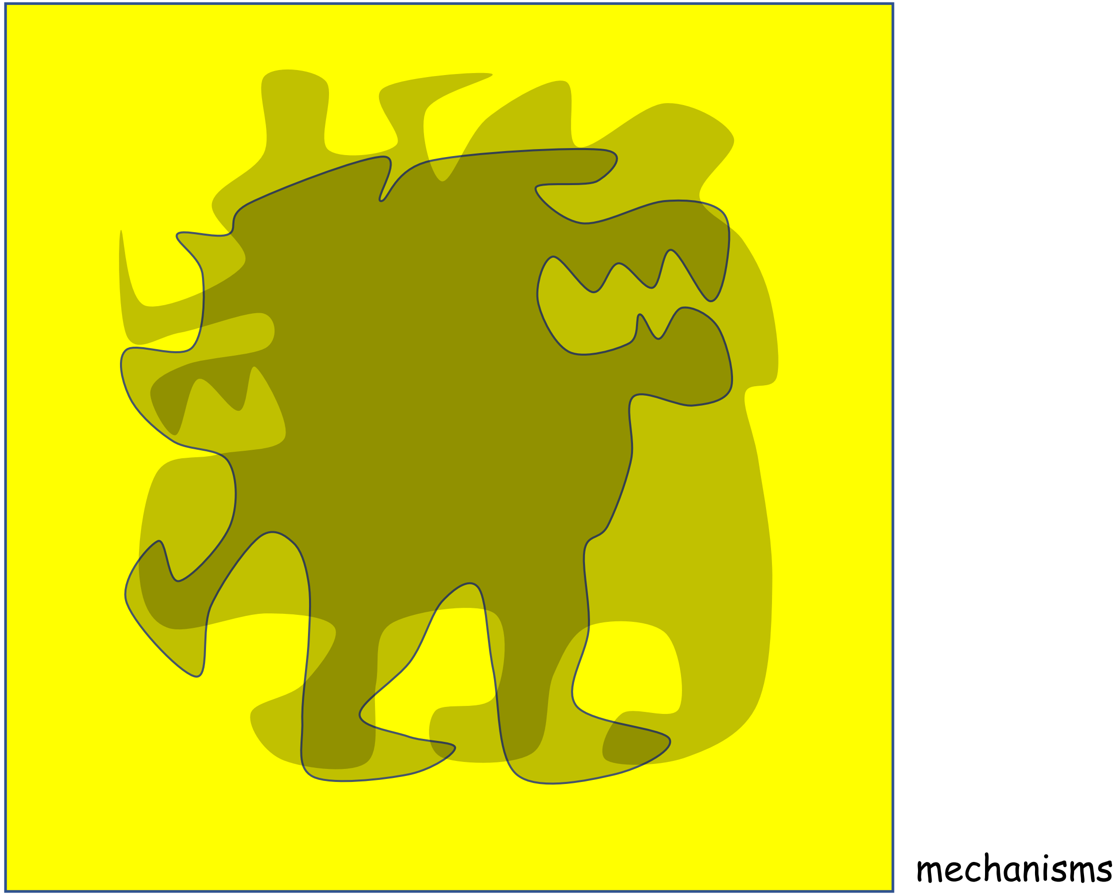
- How can I design systems without having a clear definition of what I am supposed to do?
- What if we consider all possible interpretations?



A CS interpretation of a legal privacy standard

- How can I design systems without having a clear definition of what I am supposed to do?

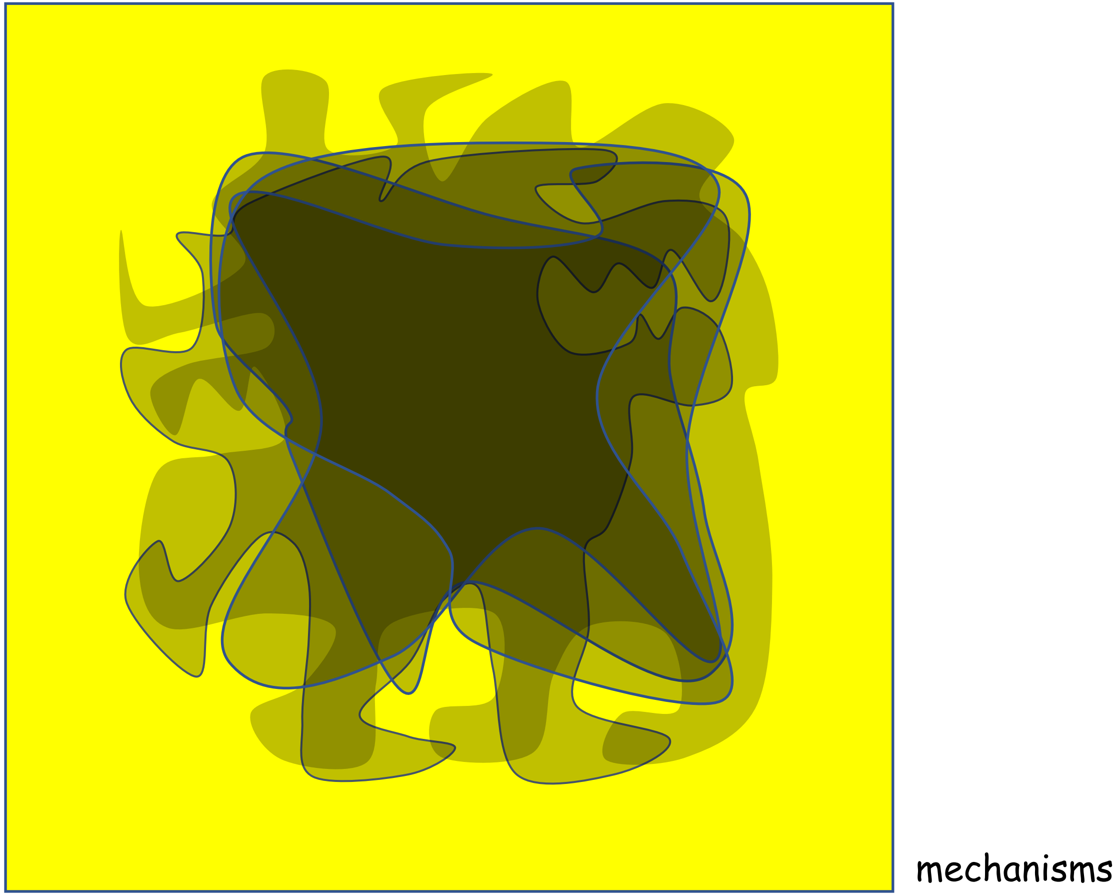
- What if we consider all possible interpretations?



A CS interpretation of a legal privacy standard

- How can I design systems without having a clear definition of what I am supposed to do?

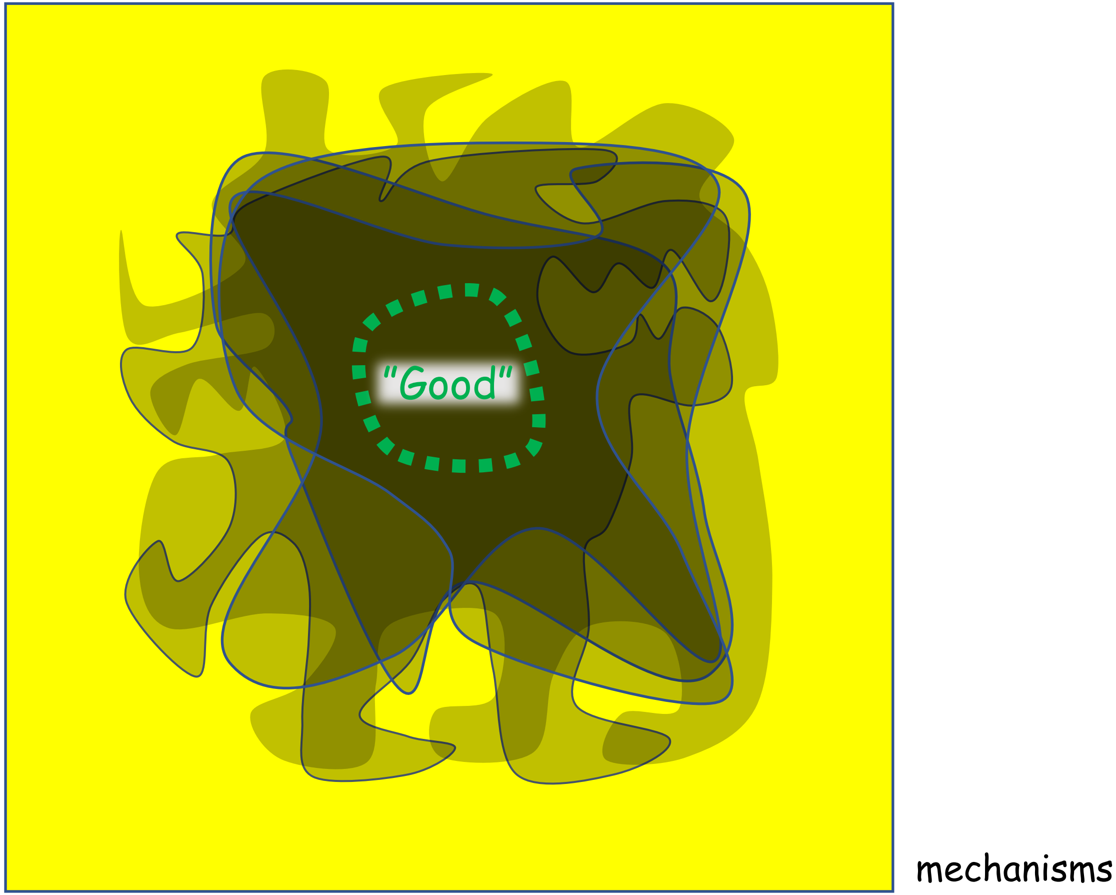
- What if we consider all possible interpretations?



A CS interpretation of a legal privacy standard

- How can I design systems without having a clear definition of what I am supposed to do?

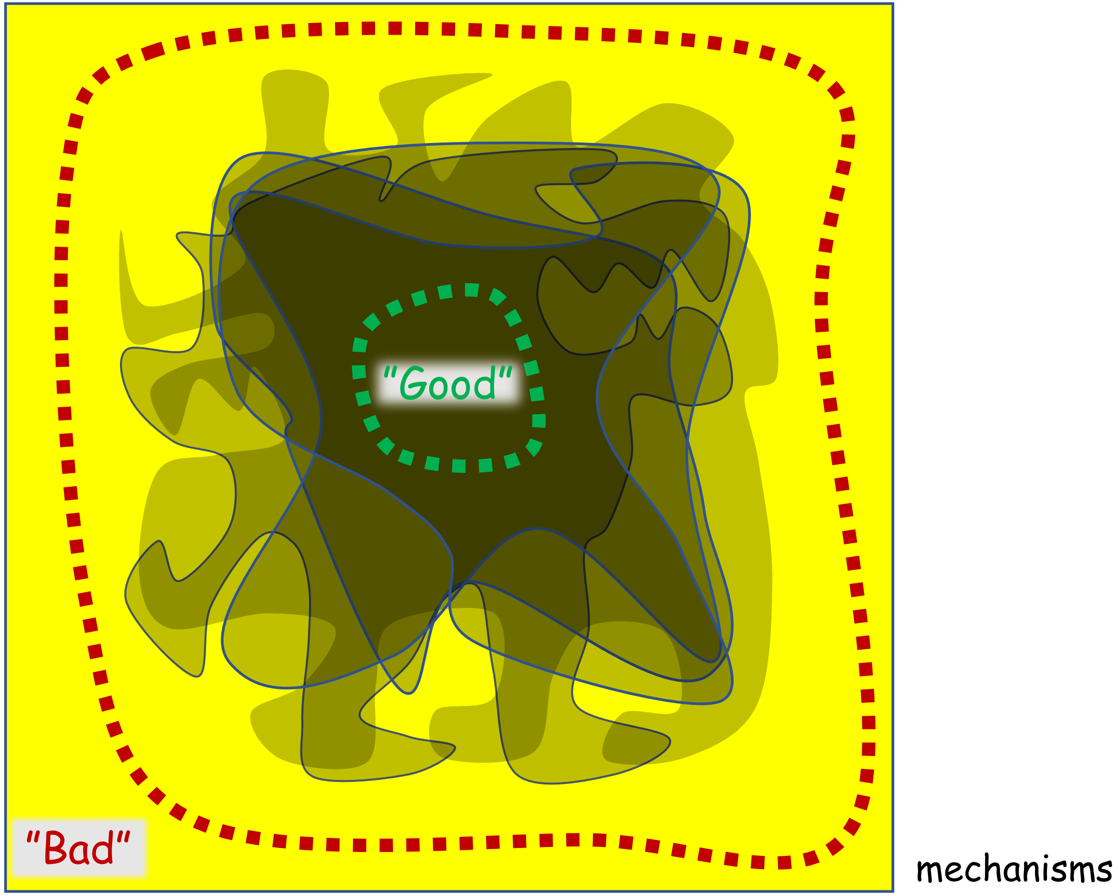
- What if we consider all possible interpretations?



A CS interpretation of a legal privacy standard

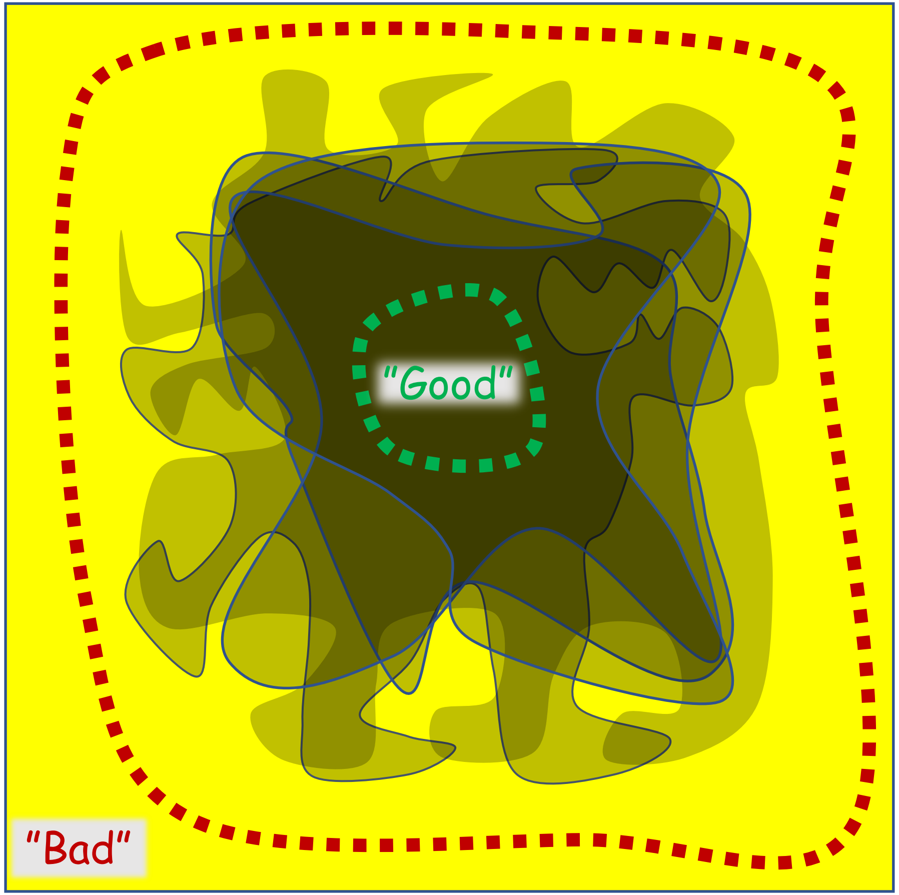
- How can I design systems without having a clear definition of what I am supposed to do?

- What if we consider all possible interpretations?



A CS interpretation of a legal privacy standard

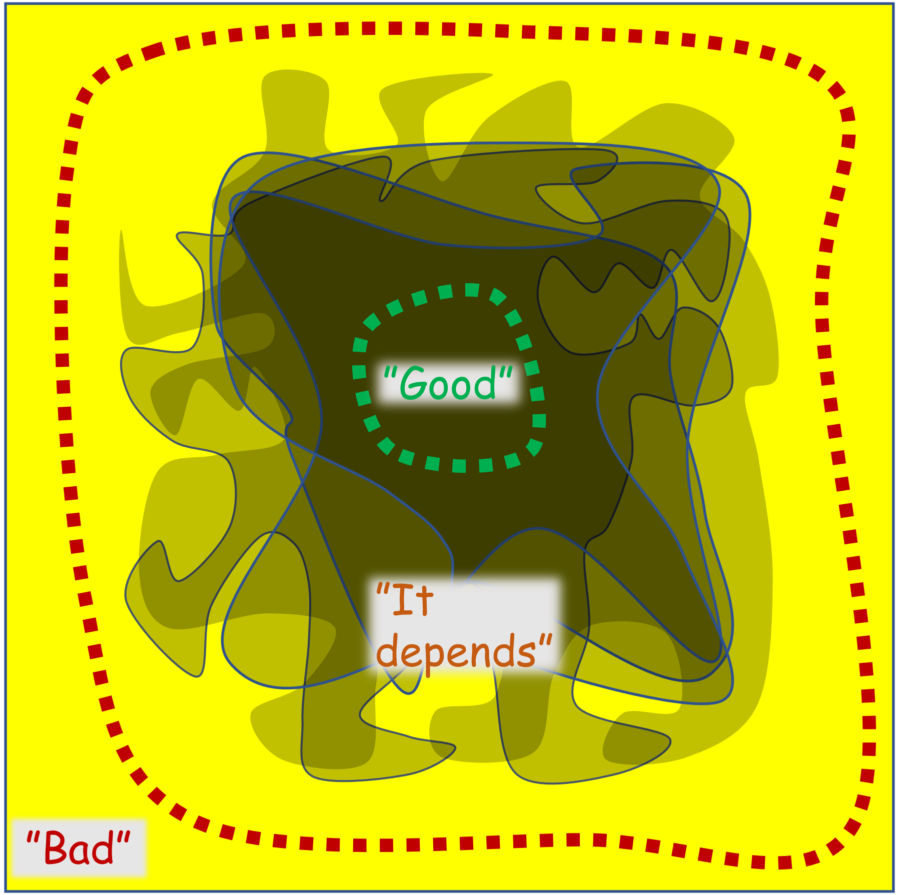
- How can I design systems without having a clear definition of what I am supposed to do?
- What if we consider all possible interpretations?
- Well defined boundaries are helpful!



mechanisms

A CS interpretation of a legal privacy standard

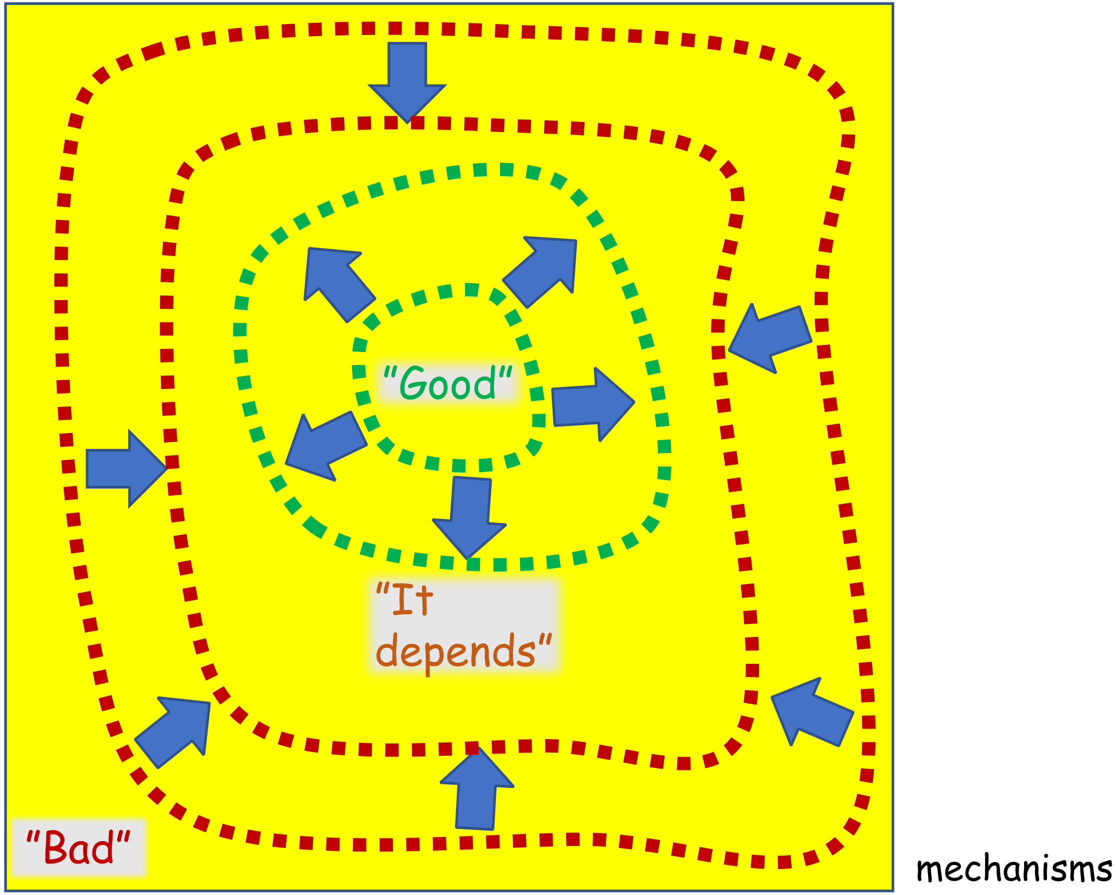
- How can I design systems without having a clear definition of what I am supposed to do?
- What if we consider all possible interpretations?
- Well defined boundaries are helpful!



mechanisms

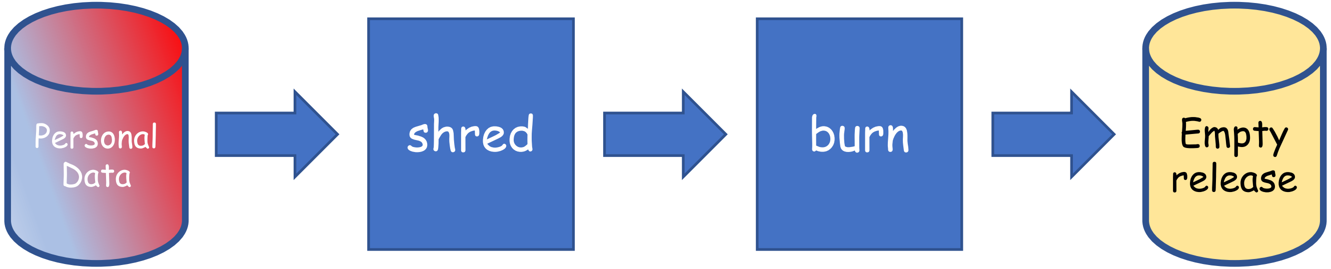
A CS interpretation of a legal privacy standard

- How can I design systems without having a clear definition of what I am supposed to do?
- What if we consider all possible interpretations?
- Well defined boundaries are helpful!
- Boundaries may become tighter as we improve our analysis



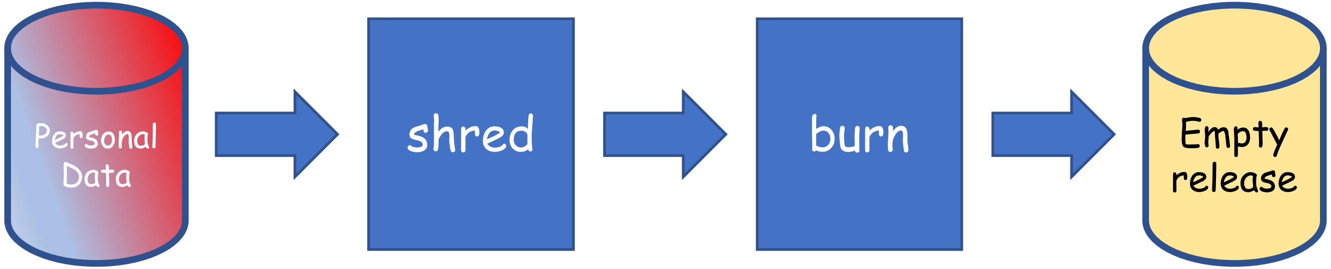
Two simple mechanism families

The empty release mechanism



- Maybe not a good use of taxpayer money
- But always protects privacy/anonymity/prevents identification

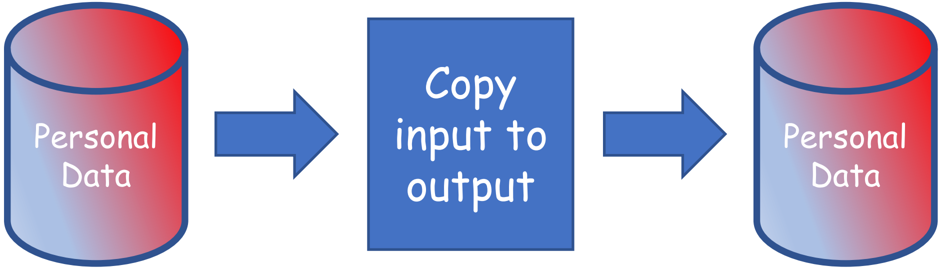
The empty release mechanism



- Maybe not a good use of taxpayer money
- But always protects privacy/anonymity/prevents identification

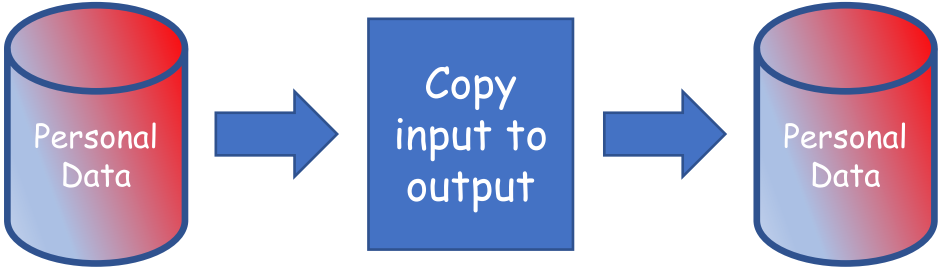
- More mechanisms of this family: any mechanism that results from postprocessing the empty release
    - Postprocessing: further processing of the outcome without looking at the data
    - E.g. the mech that ignores its data and outputs “5”

The identity mechanism



- Never protects privacy/anonymity/prevents identification

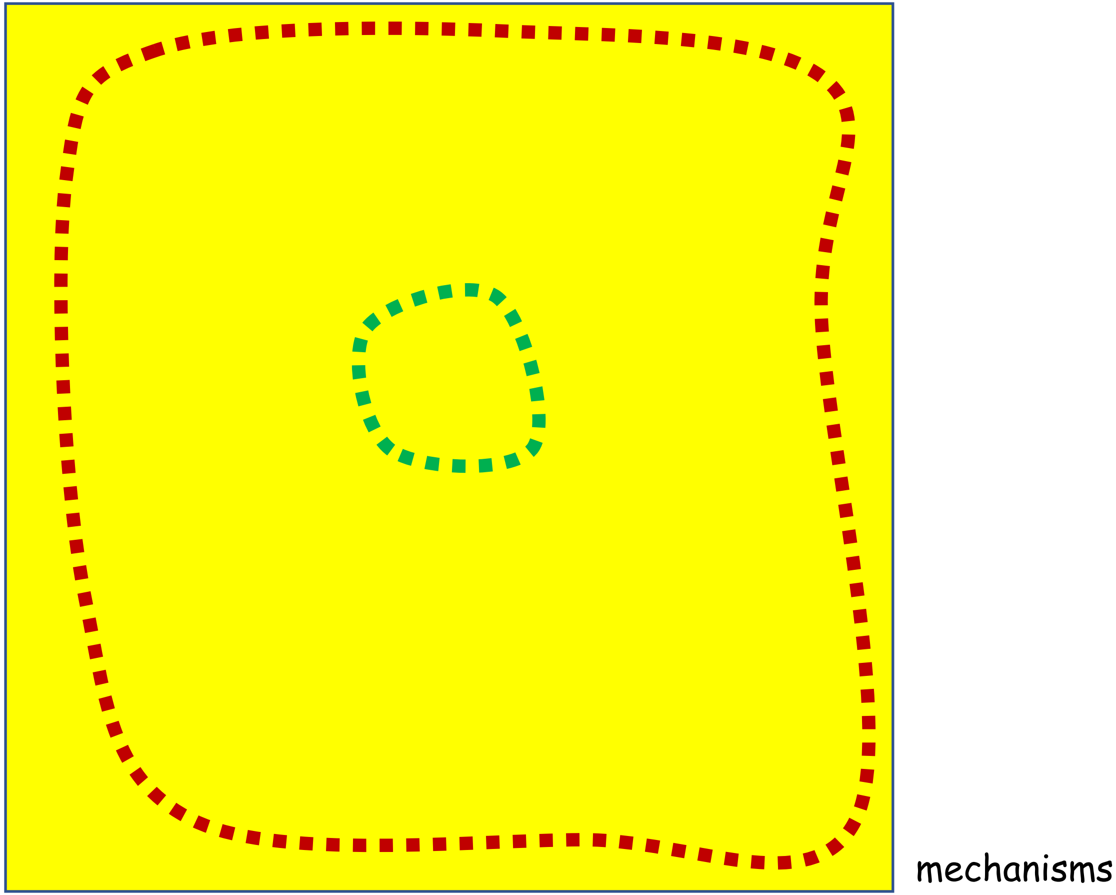
The identity mechanism



- Never protects privacy/anonymity/prevents identification

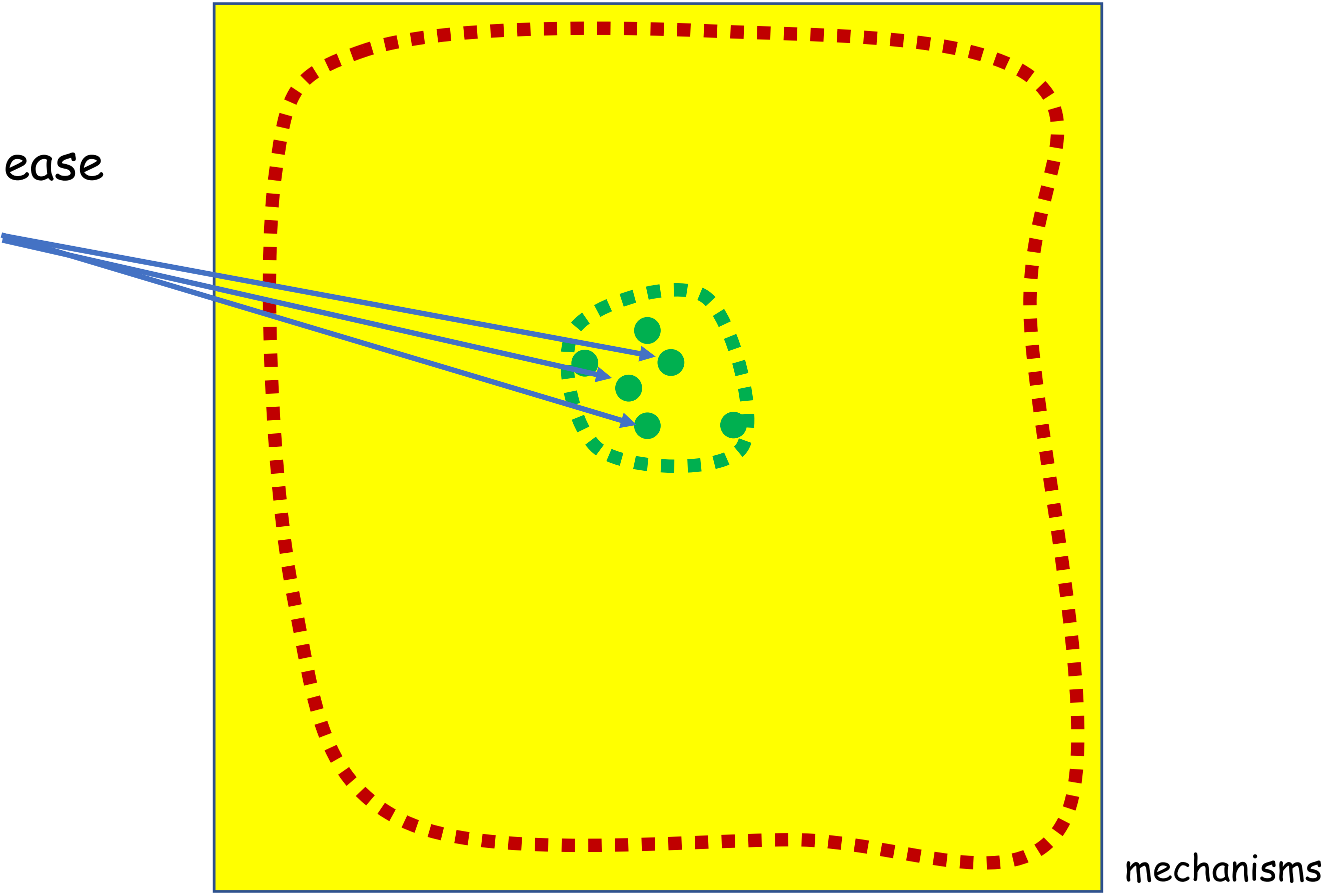
- More mechanisms in this family: any mechanism whose output can be post-processed to result in identity
    - Aka reconstruction attacks [DN03]

A CS interpretation of a legal privacy standard



A CS interpretation of a legal privacy standard

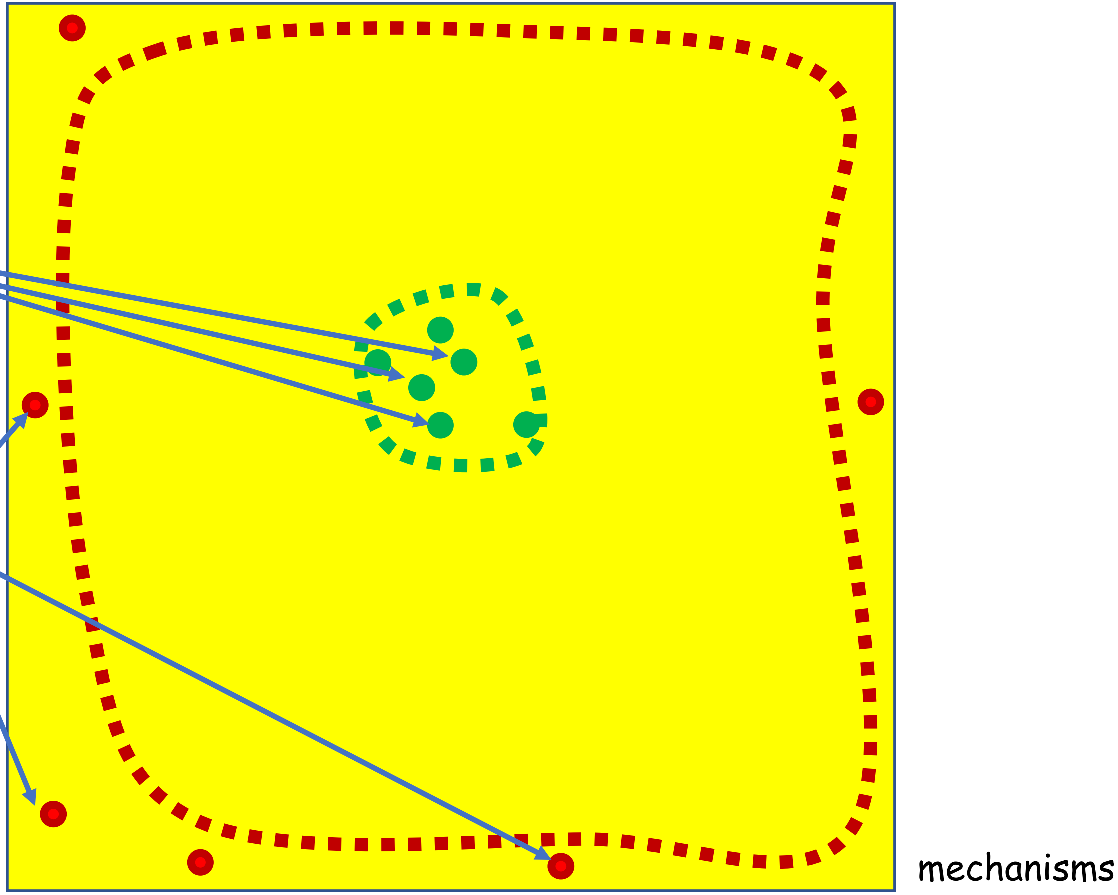
Any privacy regulation should deem empty release as privacy preserving



A CS interpretation of a legal privacy standard

Any privacy regulation should deem empty release as privacy preserving

Any privacy regulation should deem identity as not privacy preserving



How is this useful? Examples of gaining certainty

- The GDPR concept of singling out:
  - Mathematical analysis showing that the formal interpretation of the concept by A29WG excludes empty release
  - A new mathematical concept – predicate singling out – weaker requirement than that intended by the regulation
  - A “legal theorem” showing that k-anonymity does not protect against predicate singling out, and hence against the GDPR notion of singling out

How is this useful? Examples of gaining certainty

- The GDPR concept of singling out:
  - Mathematical analysis showing that the formal interpretation of the concept by A29WG excludes empty release
  - A new mathematical concept – predicate singling out – weaker requirement than that intended by the regulation
  - A "legal theorem" showing that k-anonymity does not protect against predicate singling out, and hence against the GDPR notion of singling out
- Use of differential privacy satisfies FERPA (Family Educational Rights and Privacy Act):
  - A mathematical model capturing a stronger requirement than in FERPA
  - A mathematical proof that the use of differential privacy satisfies the modeled requirements, and hence those of FERPA

Summary - towards designing systems that meet privacy laws

- Major issue: type mismatch between legal and technical definitions

Summary – towards designing systems that meet privacy laws

- Major issue: type mismatch between legal and technical definitions
- This is not about eliminating flexibility in Law or about mechanizing law.
  - Embrace flexibility but do not compromising mathematical rigor

Summary – towards designing systems that meet privacy laws

- Major issue: type mismatch between legal and technical definitions
- This is not about eliminating flexibility in Law or about mechanizing law.
  - Embrace flexibility but do not compromising mathematical rigor
- Strategy: (tightly) envelope the legal concept with mathematically defined technical concepts
  - Leads to a better understanding of privacy regulations and their interpretation
  - Leads to new technical concepts that can be used for specifying requirements and for reasoning about the adequacy of methods taken for satisfying legal requirements
  - Will hopefully lead to the development of appropriate programming tools, such as verification tools

References for this presentation

- Is privacy *privacy?* [N, Wood 2018]
- Bridging the gap between computer science and legal approaches to privacy [N, Bembenek, Wood, Bun, Gaboardi, Gasser, O'Brien, Steinke, Vadhan 2018]
- Towards formalizing the GDPR's notion of singling out [Cohen, N 2020]
- What a hybrid legal-technical analysis teaches us about privacy regulation: The case of singling out [Altman, Cohen, N, Wood 2021]
- *A Principled Approach to Defining Anonymization As Applied to EU Data Protection Law (draft)* [Altman, Cohen, Falzon, Markatou, N, Reymond, Saraogi, Wood, 2022]