# Privacy
# Principles, Properties, and Mechanisms

Anupam Datta, CMU
Michael Carl Tschantz, ICSI
Jeannette M. Wing, MSR

# Principles:
# Consumer Privacy Bill of Rights

- Individual Control
- Focused Collection
- Respect for Context

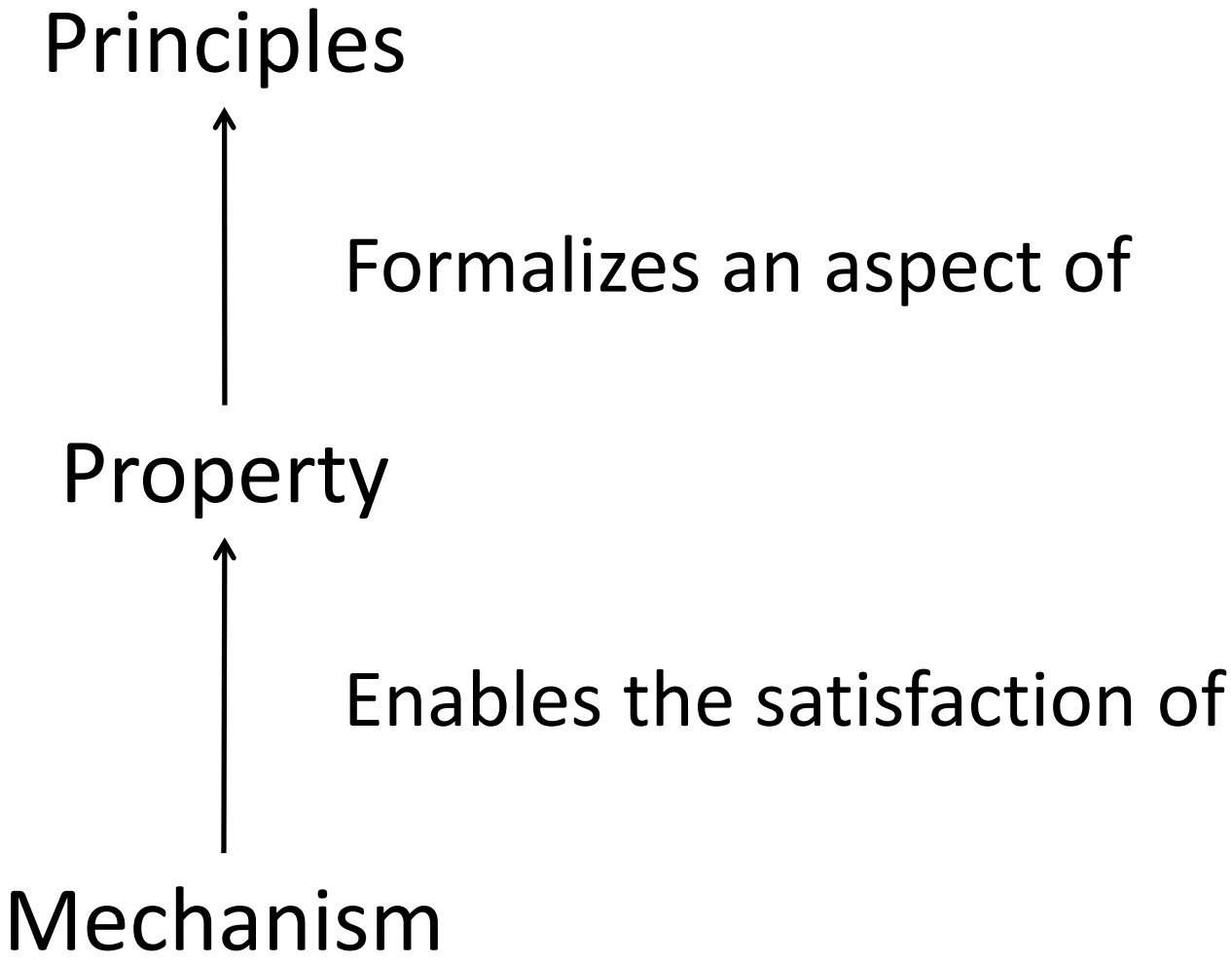
Limit information flows

- Transparency
- Access and Accuracy

Require information flows

- Security
- Accountability

Ensure the above

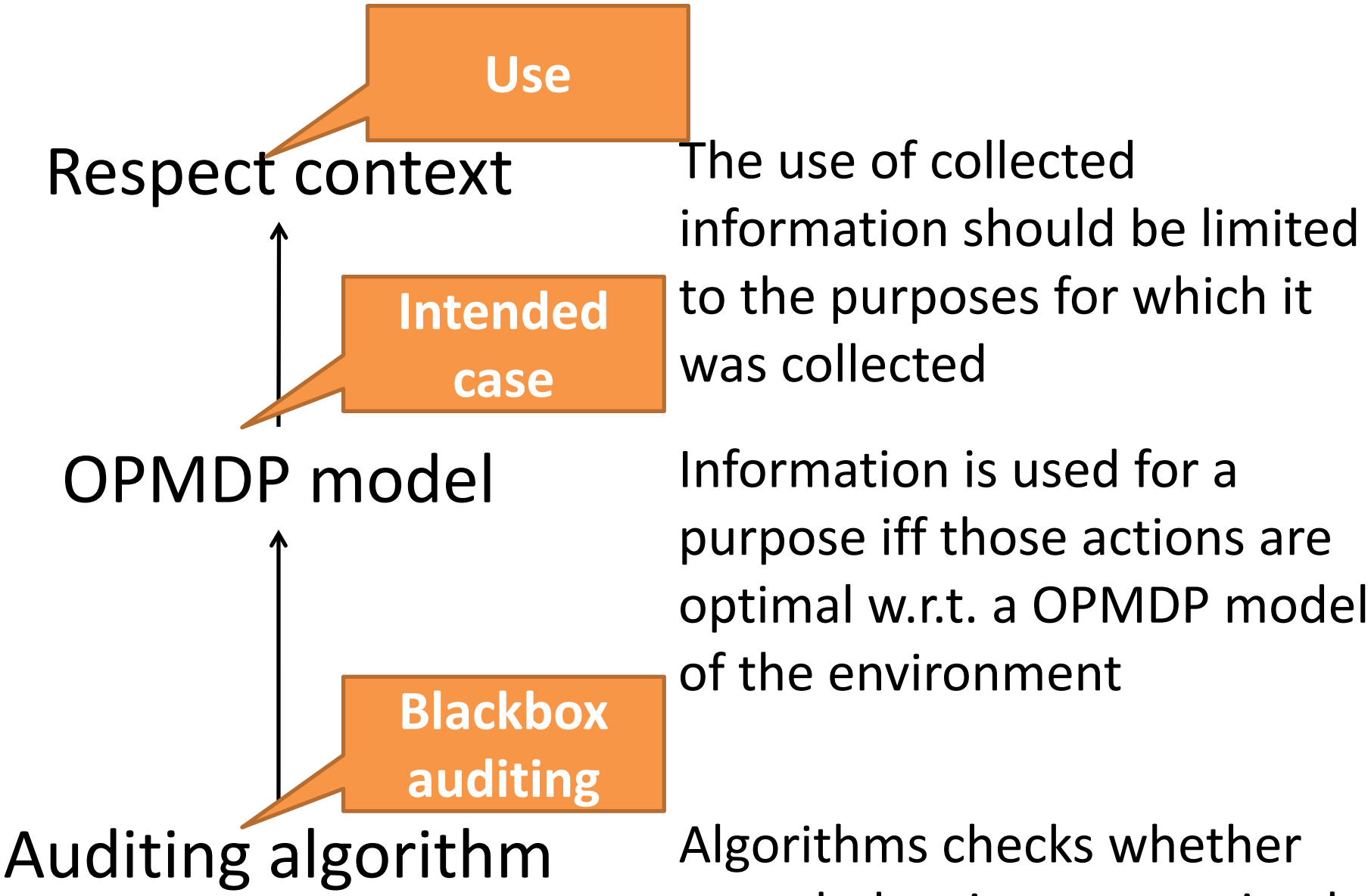

Principles
Formalizes an aspect of
Property
Enables the satisfaction of
Mechanism

| Principle | Property | Mechanism |
|---|---|---|
| **Individual Control** | Stochastic Privacy | Optimization algorithm |
| | Useable control | Inferring Interests |
| | Opting out of collection | Do Not Track |
| **Focused Collection** | Differential Privacy | Adding noise |
| | Data confidentiality | Homomorphic Encryption |
| | Allow targeting w/o collection | Tools for ads on local computer |
| | Simulatability | Privacy-Preserving Datamining |
| | Simulatability | Multi-party computation |
| | Query unknown | Private information retrieval |
| | Query unclear | TrackMeNot |
| **Respect for Context** | Traces for Contextual Integrity | Auditing algorithm |
| | Purpose OPMDP model | Auditing algorithm |
| **Transparency** | Sharing retention policy | P3P |
| | Fairness though Awareness | Distribution comparison |
| **Access and Accuracy** | Inverse Privacy | Record keeping tools |
| **Security** | Noninterference | Program analysis |
| **Accountability** | Probabilistic Noninterference | Information Flow Experiments |
| | Do not track *X* | Studies checking for tracking *X* |
| | Anonymity | De-anonymization illustrations |
| | Obey policy with judgment call | Auditing with human help |

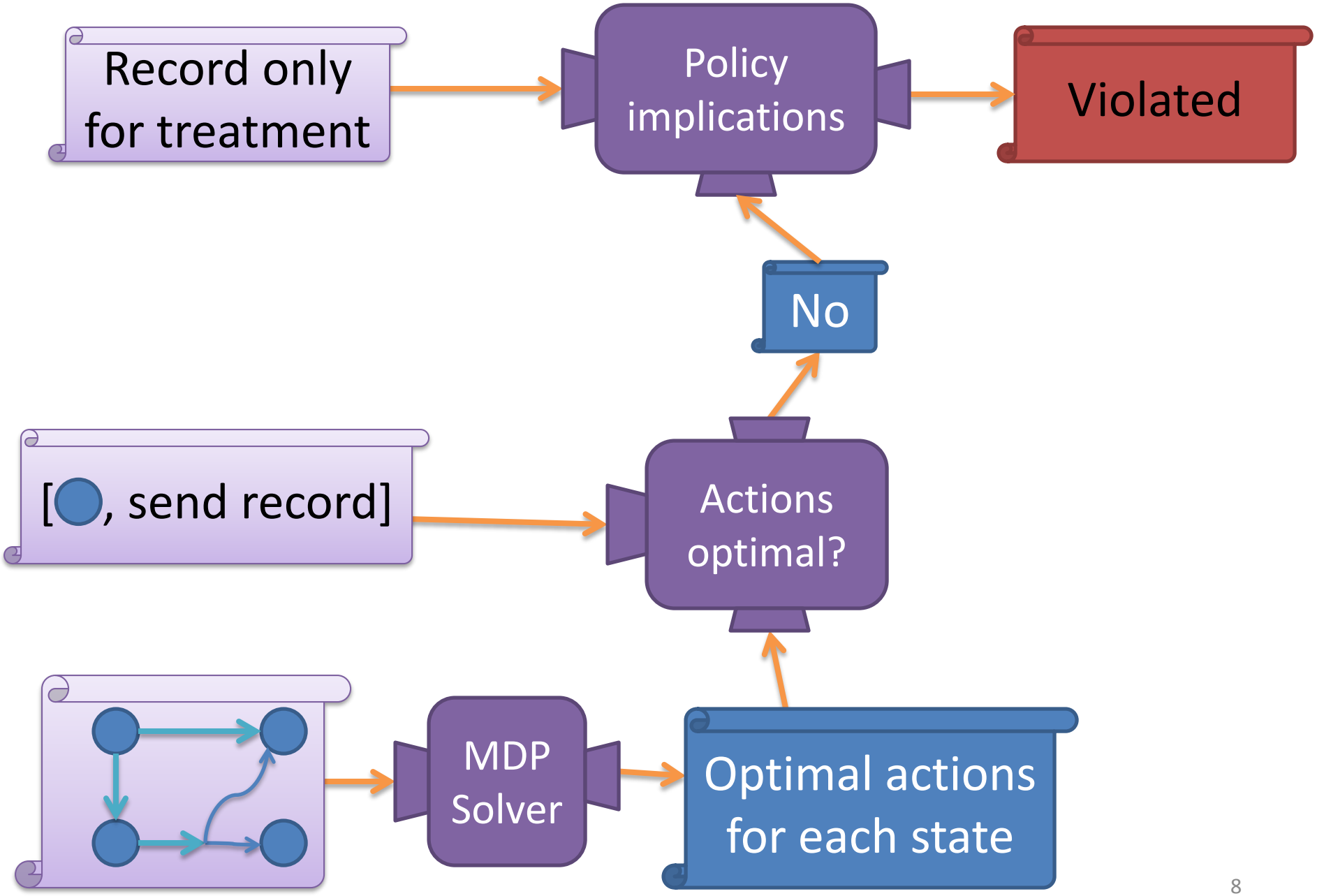
| Principle | Property | Mechanism |
|---|---|---|
| **Individual Control** | Stochastic Privacy | Optimization algorithm |
| | Useable control | Inferring Interests |
| | Opting out of collection | Do Not Track |
| **Focused Collection** | Differential Privacy | Adding noise |
| | Data confidentiality | Homomorphic Encryption |
| | Allow targeting w/o collection | Tools for ads on local computer |
| | Simulatability | Privacy-Preserving Datamining |
| | Simulatability | Multi-party computation |
| | Query unknown | Private information retrieval |
| | Query unclear | TrackMeNot |
| **Respect for Context** | Traces for Contextual Integrity | Auditing algorithm |
| | Purpose OPMDP model | Auditing algorithm |
| **Transparency** | Sharing retention policy | P3P |
| | Fairness though Awareness | Distribution comparison |
| **Access and Accuracy** | Inverse Privacy | Record keeping tools |
| **Security** | Noninterference | Program analysis |
| **Accountability** | Probabilistic Noninterference | Information Flow Experiments |
| | Do not track *X* | Studies checking for tracking *X* |
| | Anonymity | De-anonymization illustrations |
| | Obey policy with judgment call | Auditing with human help |

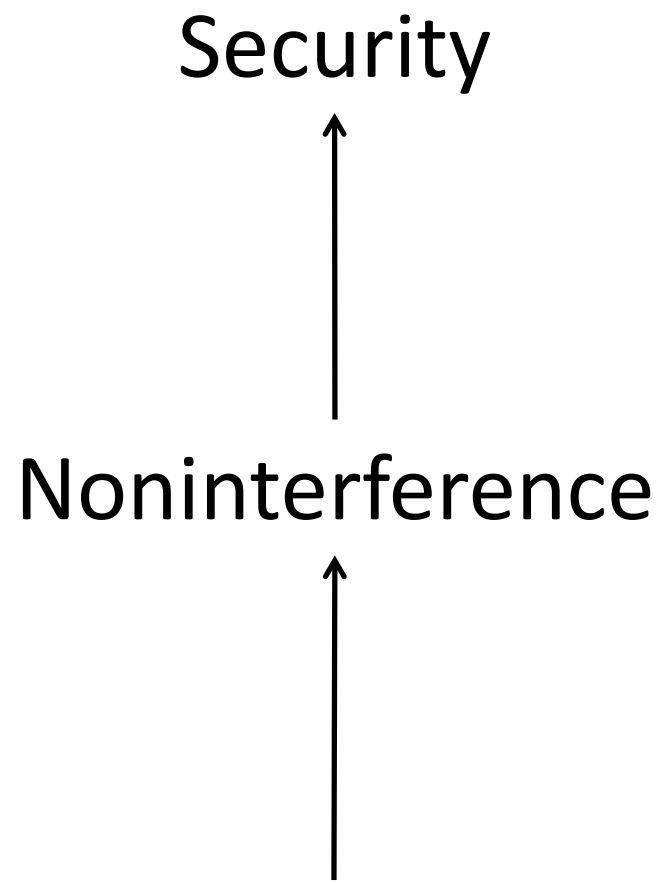
# Why Privacy Research is Interesting

- Balancing tradeoffs between
  - data holder
  - data subjects
  - public
- Adversarial vs. differing goals
- Intended case interesting
- Quantitative properties
- External blackbox methods
  - Auditing
  - Experimentation
- Use


Use
Respect context
The use of collected information should be limited to the purposes for which it was collected
Intended case
OPMDP model
Information is used for a purpose iff those actions are optimal w.r.t. a OPMDP model of the environment
Blackbox auditing
Auditing algorithm
Algorithms checks whether recorded actions are optimal

[Tschantz, Datta, Wing 2012, 2013]






Sensitive information shouldn’t flow to public outputs

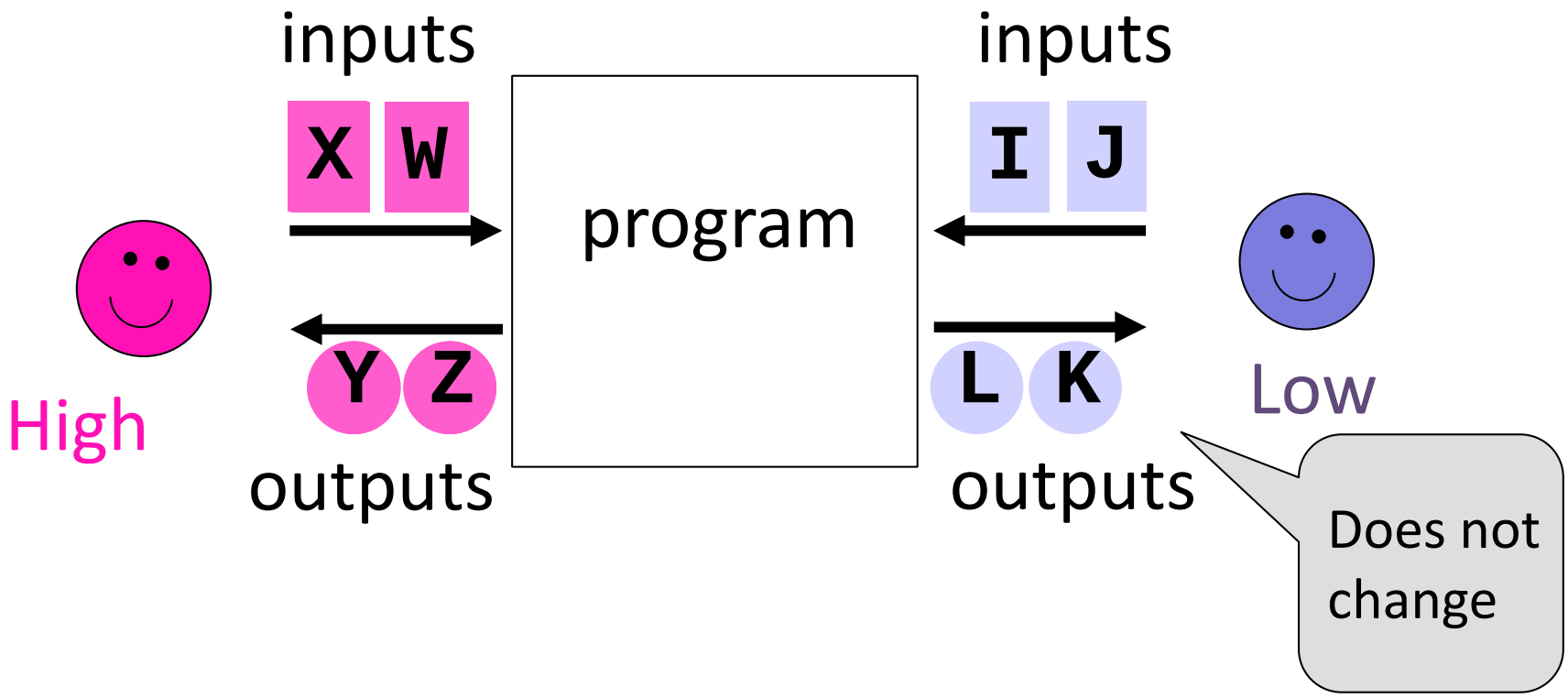
The low-level outputs look the same under any two high-level inputs

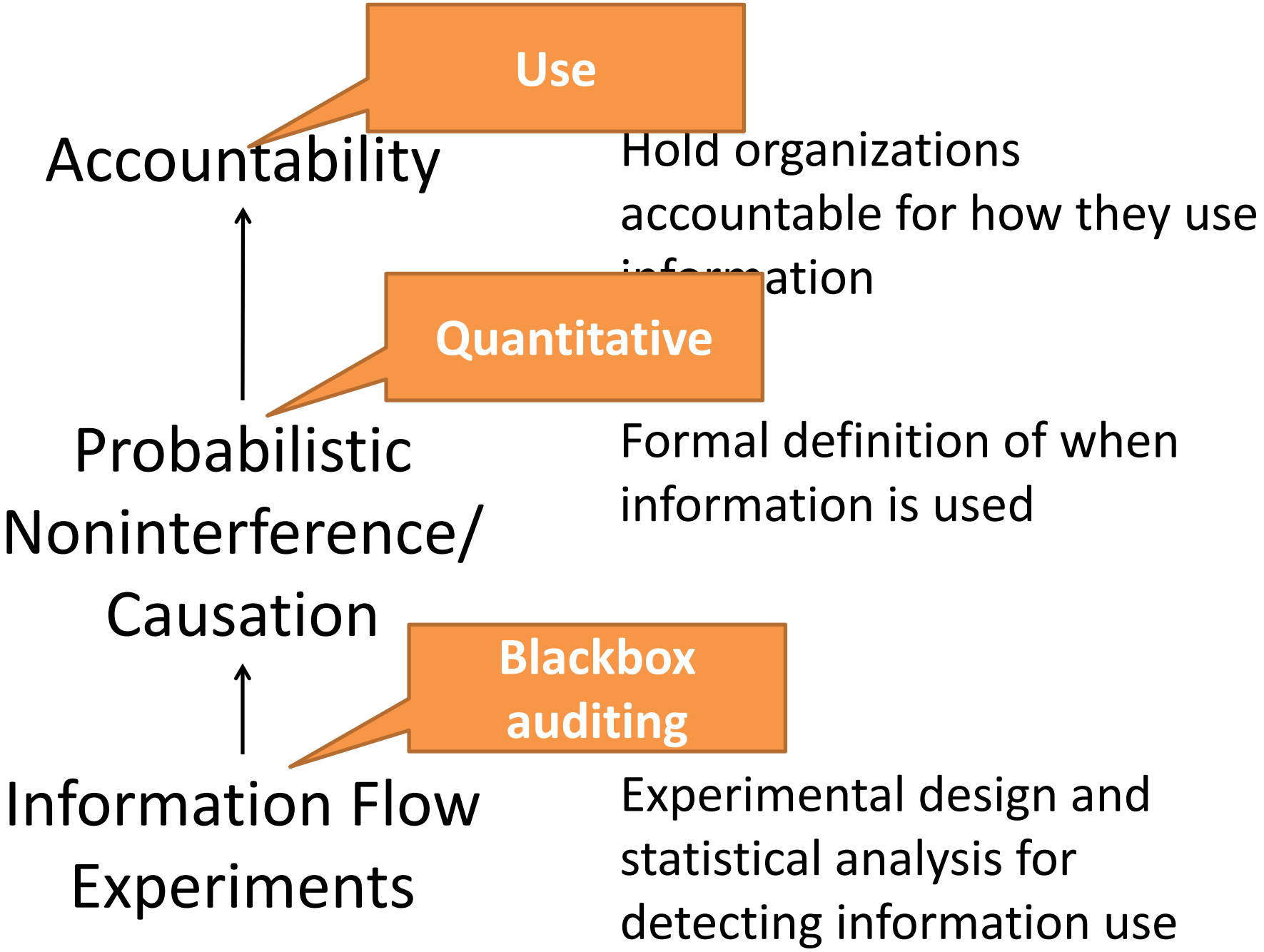
Checked programs have noninterference

[Goguen and Meseguer, 1982]

# Noninterference

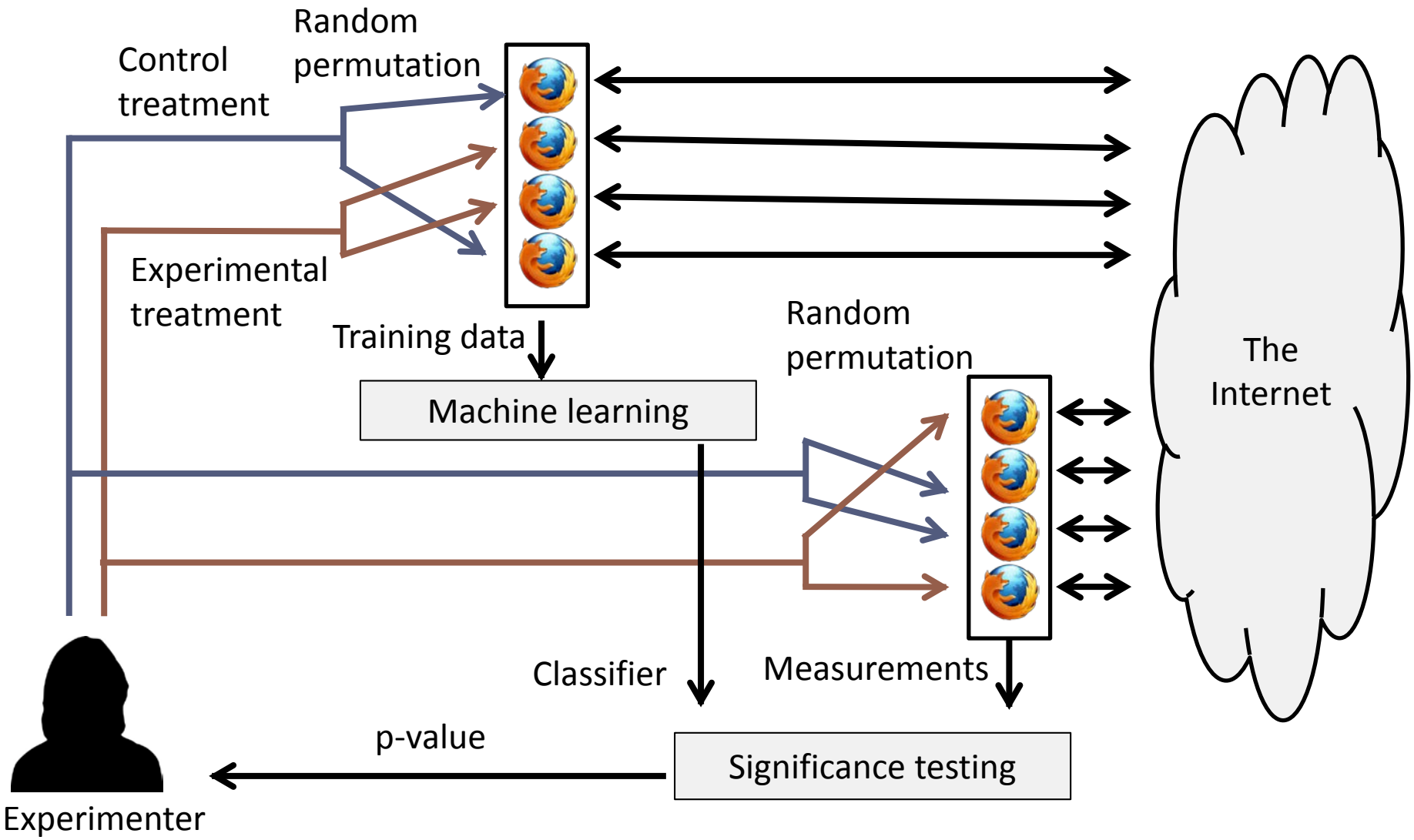
Rule out flows from High inputs to Low outputs



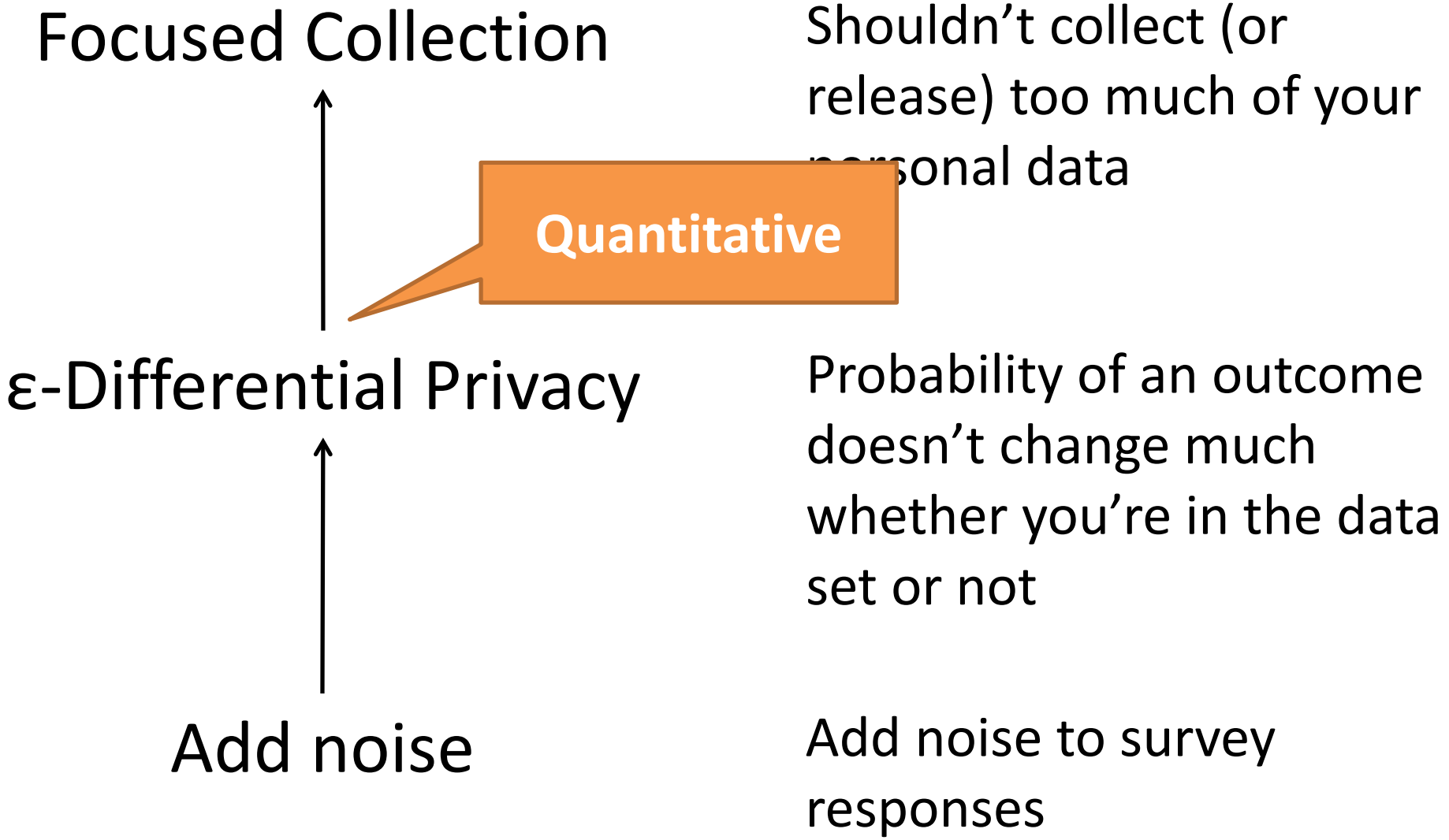

Use
Accountability
Hold organizations accountable for how they use information
Quantitative
Probabilistic Noninterference/ Causation
Formal definition of when information is used
Blackbox auditing
Information Flow Experiments
Experimental design and statistical analysis for detecting information use

[Datta, Tschantz, Datta & Wing, 2014]

# Information Flow Experiments

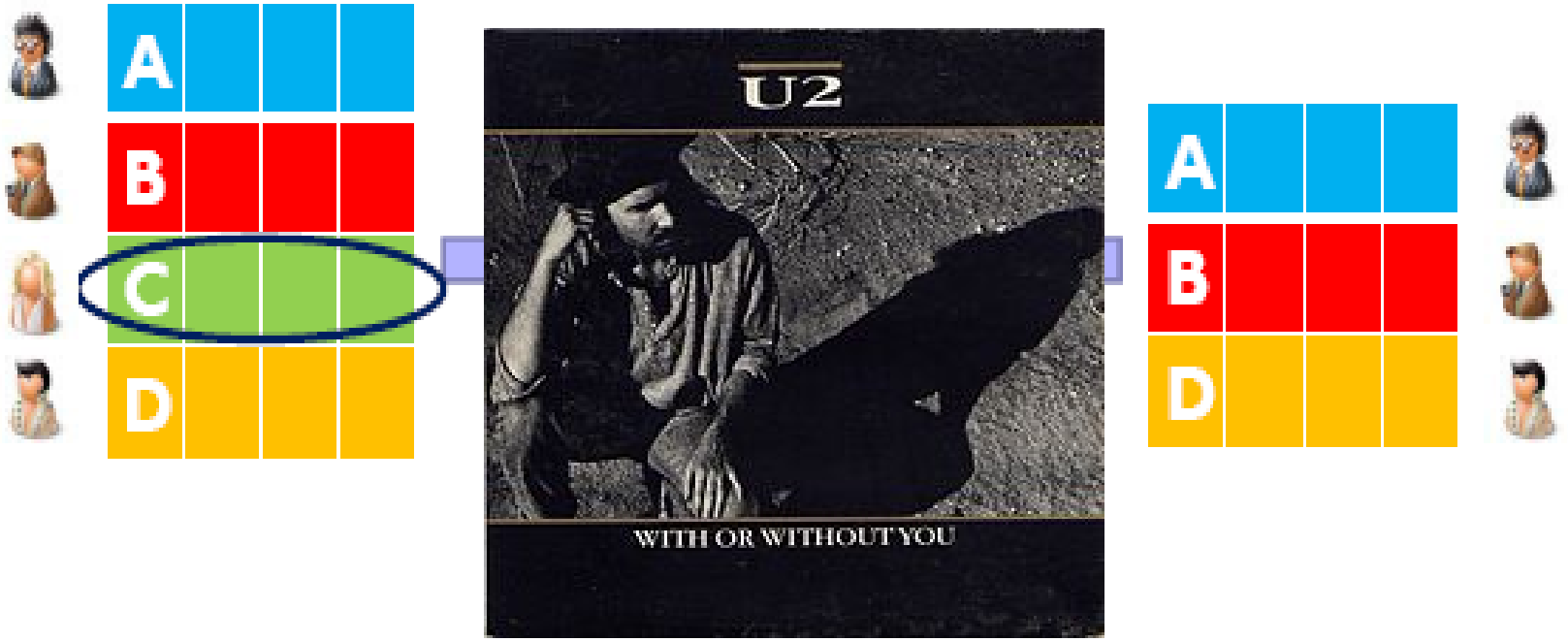


| Principles | Property | Mechanism |
|---|---|---|
| **Individual Control** | Stochastic Privacy | Optimization algorithm |
| | Useable control | Tags |
| **Focused Collection** | Differential Privacy | Adding noise |
| | Data confidentiality | Homomorphic Encryption |
| | Allow targeting w/o collection | Tools for ads on local computer |
| | Simulatability | Privacy-Preserving Datamining |
| | Simulatability | Multi-party computation |
| | Query unknown | Private information retrieval |
| **Respect for Context** | Traces for Contextual Integrity | Auditing algorithm |
| | Purpose OPMDP model | Auditing algorithm |
| **Transparency** | Sharing retention policy | P3P |
| | Fairness though Awareness | Distribution comparison |
| **Access and Accuracy** | Inverse Privacy | Record keeping tools |
| **Security** | Noninterference | Program analysis |
| **Accountability** | Probabilistic Noninterference | Information Flow Experiments |
| | Do not track *X* | Studies checking for tracking *X* |
| | Obey policy with judgment call | Auditing with human help |

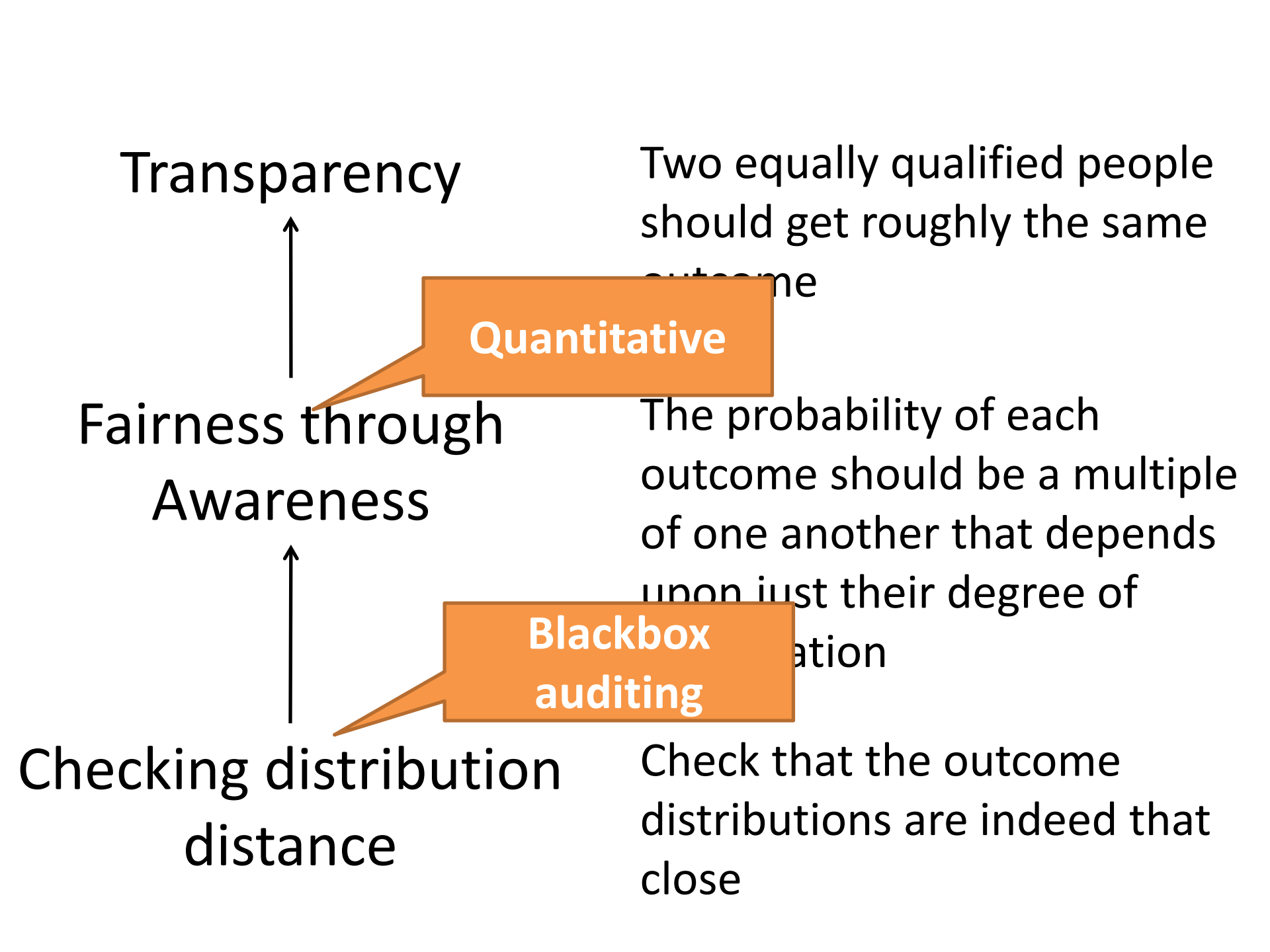

Focused Collection
Shouldn’t collect (or release) too much of your personal data
Quantitative
ε-Differential Privacy
Probability of an outcome doesn’t change much whether you’re in the data set or not
Add noise
Add noise to survey responses

# Differential Privacy: Idea

[Dwork, McSherry, Nissim, Smith 2006]




Released statistic is about the same if any individual's record is removed from the database


Transparency
Two equally qualified people should get roughly the same
Quantitative
Fairness through Awareness
The probability of each outcome should be a multiple of one another that depends
Blackbox auditing
Checking distribution distance
Check that the outcome distributions are indeed that close

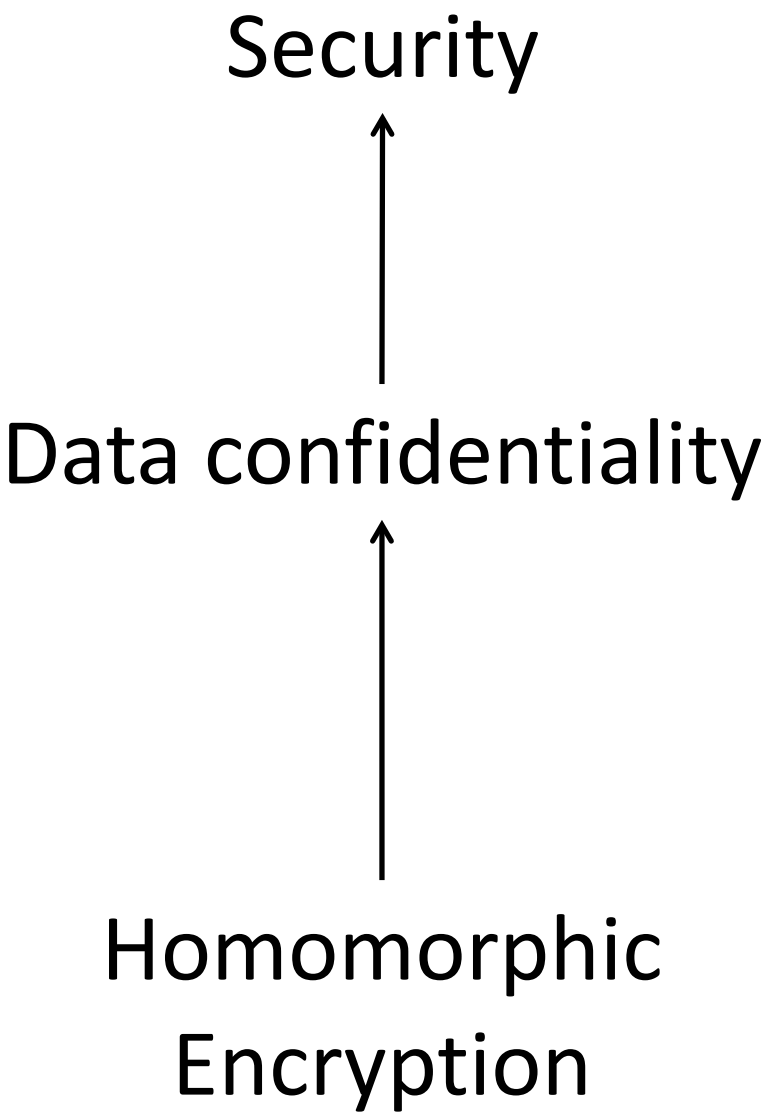
# Individual Fairness

[Dwork, Hardt, Pitassi, Reingold, Zemel 2011]

**Treat *similar* individuals *similarly***

Similar for the purpose of the classification task

Similar distribution over outcomes



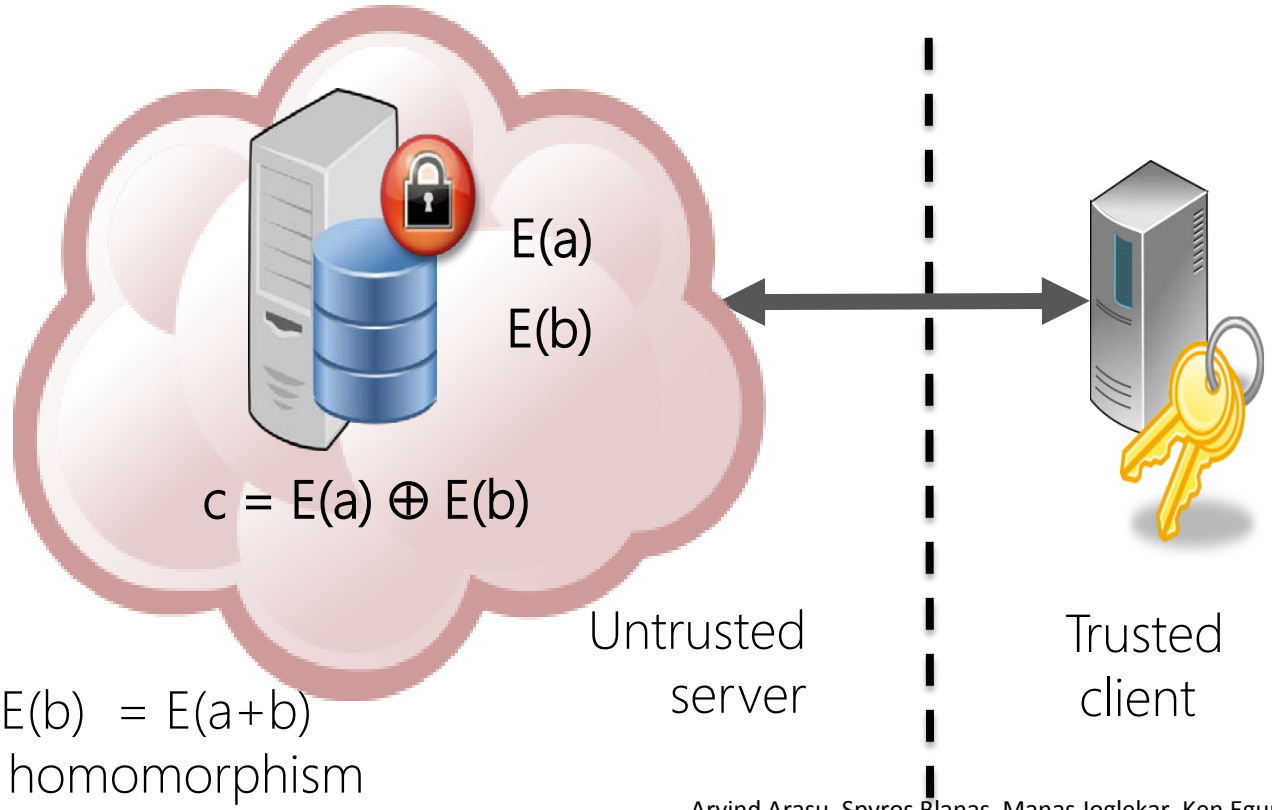

Want to keep data in the cloud without allowing cloud provider to access it

Sensitive data never appears in plaintext on the untrusted server

Allows computation over encrypted databases

# Computing over Encrypted Data

**Privacy Guarantee**: Sensitive data never appears in plaintext on the untrusted server




E(a) ⊕ E(b) = E(a+b)
due to homomorphism

Arvind Arasu, Spyros Blanas, Manas Joglekar, Ken Eguro, Raghav Kaushik, Donald Kossmann, Ravi Ramamurthy, Prasang Upadhyaya, and Ramarathnam Venkatesan, Engineering Performance and Security with Cipherbase, in *Data Engineering Bulletin*, IEEE, December 2012

# Implementations for SQL Databases

- **CryptDB (MIT)**
  - Raluca Ada Popa, Catherine M. S. Redfield, Nickolai Zeldovich, and Hari Balakrishnan, **CryptDB: Protecting Confidentiality with Encrypted Query Processing**, SOSP 2011, http://people.csail.mit.edu/nickolai/papers/raluca-cryptdb.pdf
- **Cipherbase (Microsoft Research)**
  - Arvind Arasu, Spyros Blanas, Manas Joglekar, Ken Eguro, Raghav Kaushik, Donald Kossmann, Ravi Ramamurthy, Prasang Upadhyaya, and Ramarathnam Venkatesan, Engineering Performance and Security with Cipherbase, in *Data Engineering Bulletin*, IEEE, December 2012
  - Arvind Arasu, Ken Eguro, Manas Joglekar, Raghav Kaushik, Donald Kossmann, and Ravi Ramamurthy, **Transaction Processing on Confidential Data using Cipherbase**, ICDE 2015, http://research.microsoft.com/apps/pubs/default.aspx?id=231354
- **Monomi (MIT)**
  - Stephen Tu, M. Frans Kaashoek, Samuel Madden, Nickolai Zeldovich, **Processing Analytical Queries over Encrypted Data,** PVLDB 2013.
- **TrustedDB (Stony Brook)**
  - Sumeet Bajaj, Radu Sion, **TrustedDB: A Trusted Hardware-Based Database with Privacy and Data Confidentiality**, *IEEE Transactions on Knowledge & Data Engineering*, vol.26, no. 3, pp. 752-765, March 2014, doi:10.1109/TKDE.2013.38

Focused Collection

Learn only output of computation; nothing else about secret inputs of individual parties

Simulatability

Real interaction indistinguishable from interaction involving trusted third party

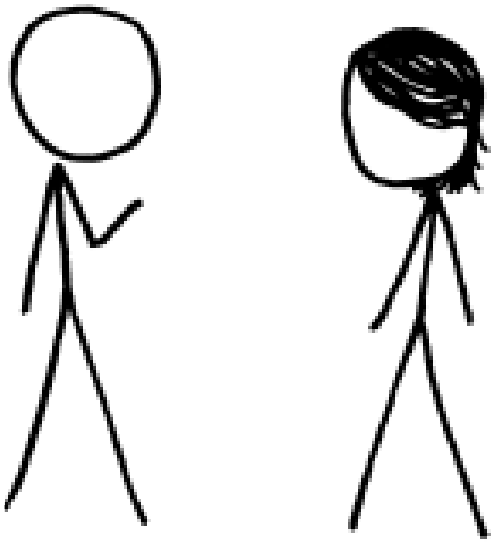
Multi-party computation

Generic protocols for any efficiently computable function [Yao82,GMW87]

Slide: David Evans et al.

# Secure Two-Party Computation

Bob's Salary: $50,000



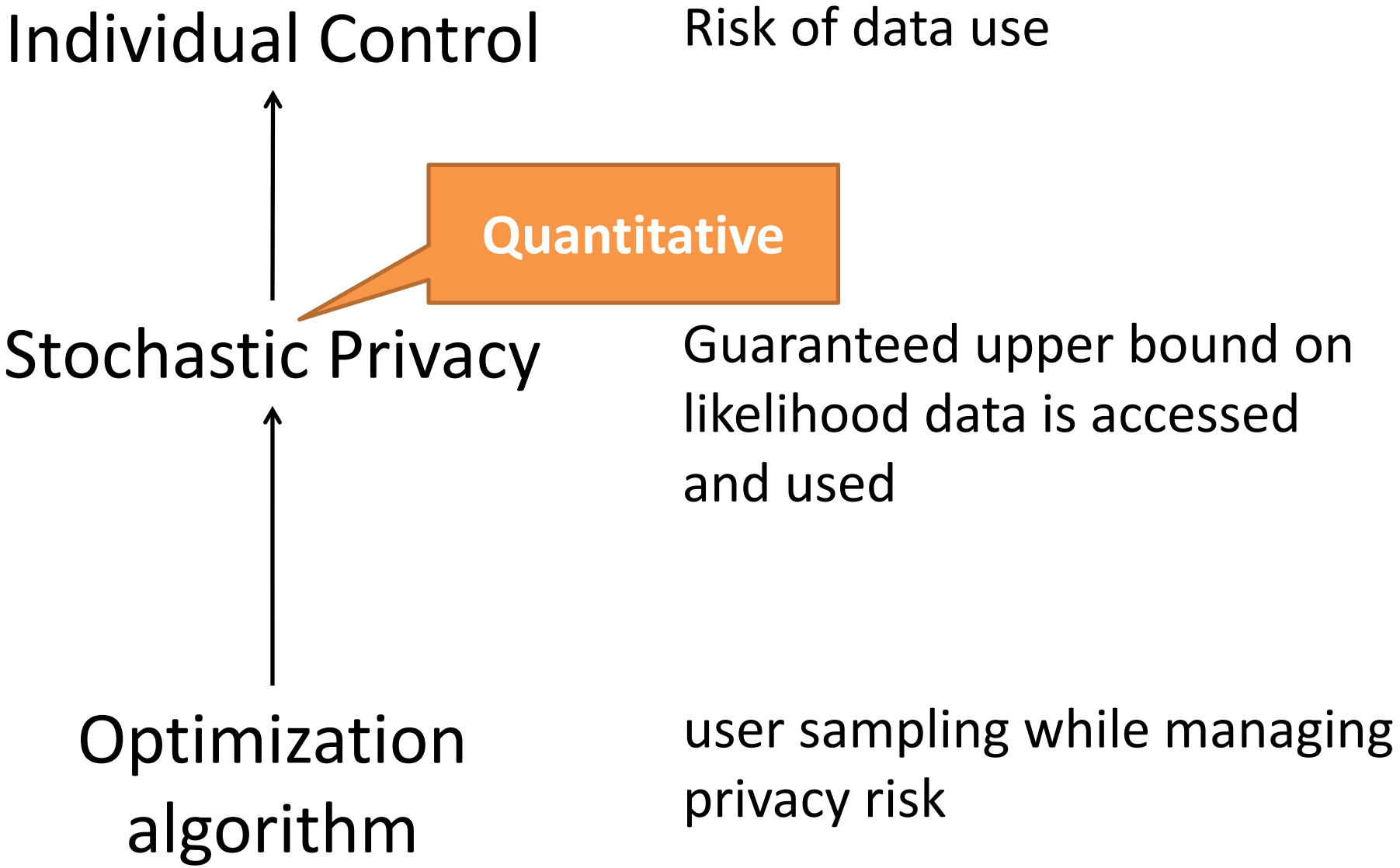
Alice's Salary: $75,000

**Bob**

**Alice**

$$x = f(g_A, g_B)$$

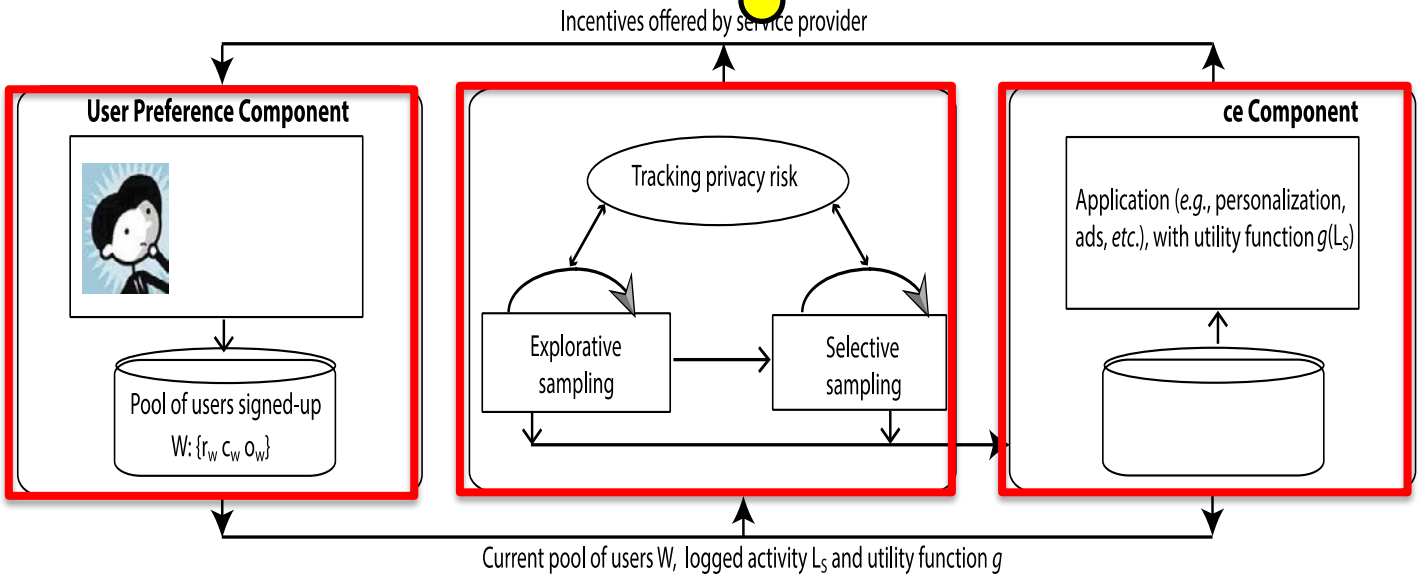
Can Alice and Bob compute a function of their private data, without exposing anything about their data besides the result?


Individual Control
Risk of data use
Quantitative
Stochastic Privacy
Guaranteed upper bound on likelihood data is accessed and used
Optimization algorithm
user sampling while managing privacy risk

# Stochastic Privacy

[Adish Singla, Eric Horvitz, Ece Kamar, Ryen White, AAAI'14]

**Privacy Property:** Guaranteed upper bound on likelihood data is accessed and used by service provider, given user stated risk preference, *r.*




- **User preferences** – choosing risk *r*, offer incentives
- **System preferences** – application utility, *g*
- **Optimization** – user sampling while managing privacy risk

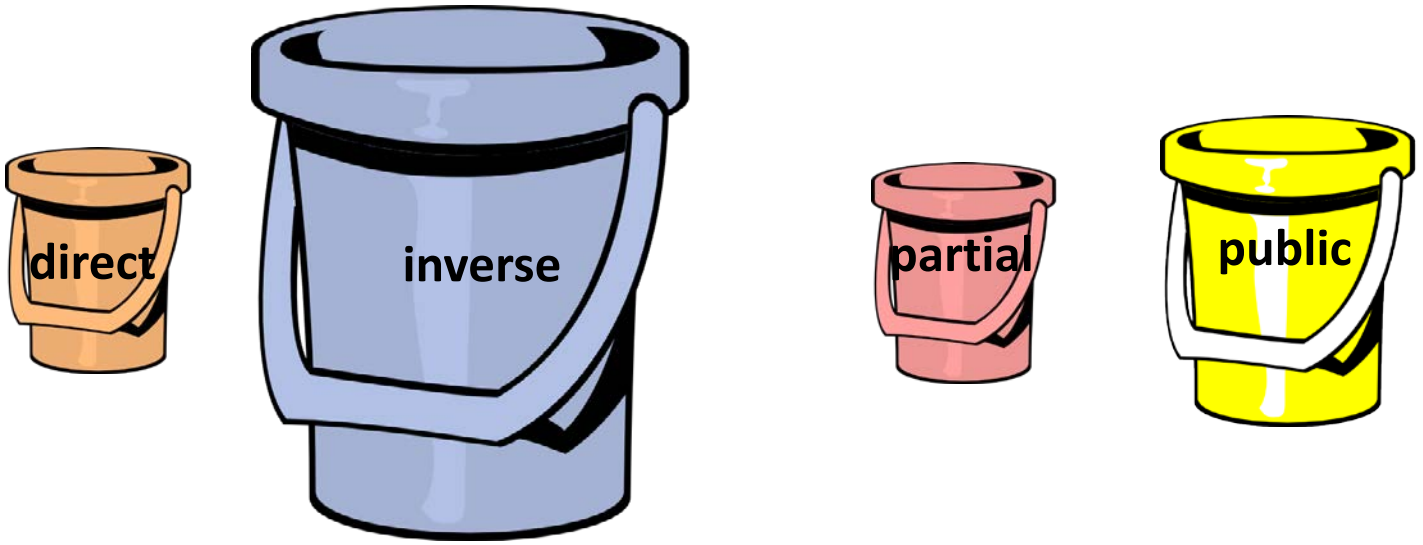
Access and accuracy

Inverse Privacy

Record keeping tools

People should know the information that companies keep on them

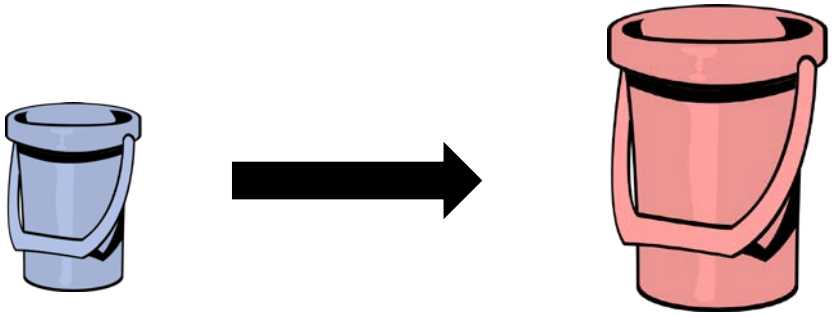
# Four-Bucket Classification of Personal Information




1. The information about you that you have and nobody else does.
   To contrast this bucket with the next one, we call it *directly private*.
2. **The *inversely private* information about you, information that some party has but you don't.**
3. The *partially private* information about you, information that you and a limited number of other parties have.
4. The *public information* about you.

[Yuri Gurevich, Efim Hudis, Jeannette Wing, MSR TR 2014-100, July 2014, http://research.microsoft.com/apps/pubs/default.aspx?id=225005]

# Proposal: Inverse to Partial Privacy



**inverse** **partial**

- **Technology:** Develop tools that enhance people's capacity to keep records.
- **Legal:** Make institutions legally responsible to share back information.
- **Economics:** Create technological, business, and social incentives to entice institutions to share information back.
- **Norms:** Encourage the creation of a new social norm, where person-to-institution interactions produce partially private information only.

| Principles | Property | Mechanism |
|---|---|---|
| **Individual Control** | Stochastic Privacy | Optimization algorithm |
| | Useable control | Tags |
| **Focused Collection** | Differential Privacy | Adding noise |
| | Data confidentiality | Homomorphic Encryption |
| | Allow targeting w/o collection | Tools for ads on local computer |
| | Simulatability | Privacy-Preserving Datamining |
| | Simulatability | Multi-party computation |
| | Query unknown | Private information retrieval |
| **Respect for Context** | Traces for Contextual Integrity | Auditing algorithm |
| | Purpose OPMDP model | Auditing algorithm |
| **Transparency** | Sharing retention policy | P3P |
| | Fairness though Awareness | Distribution comparison |
| **Access and Accuracy** | Inverse Privacy | Record keeping tools |
| **Security** | Noninterference | Program analysis |
| **Accountability** | Probabilistic Noninterference | Information Flow Experiments |
| | Do not track *X* | Studies checking for tracking *X* |
| | Obey policy with judgment call | Auditing with human help |

# Looking for more on

- Agreement on desired properties
- Mandatory vs. discretionary privacy
- Civic minded privacy
- Tighter integration with behavioral/social science work