The Fundamental Limits of Data Privacy

Peter Kairouz

ECE Department

University of Illinois at Urbana-Champaign

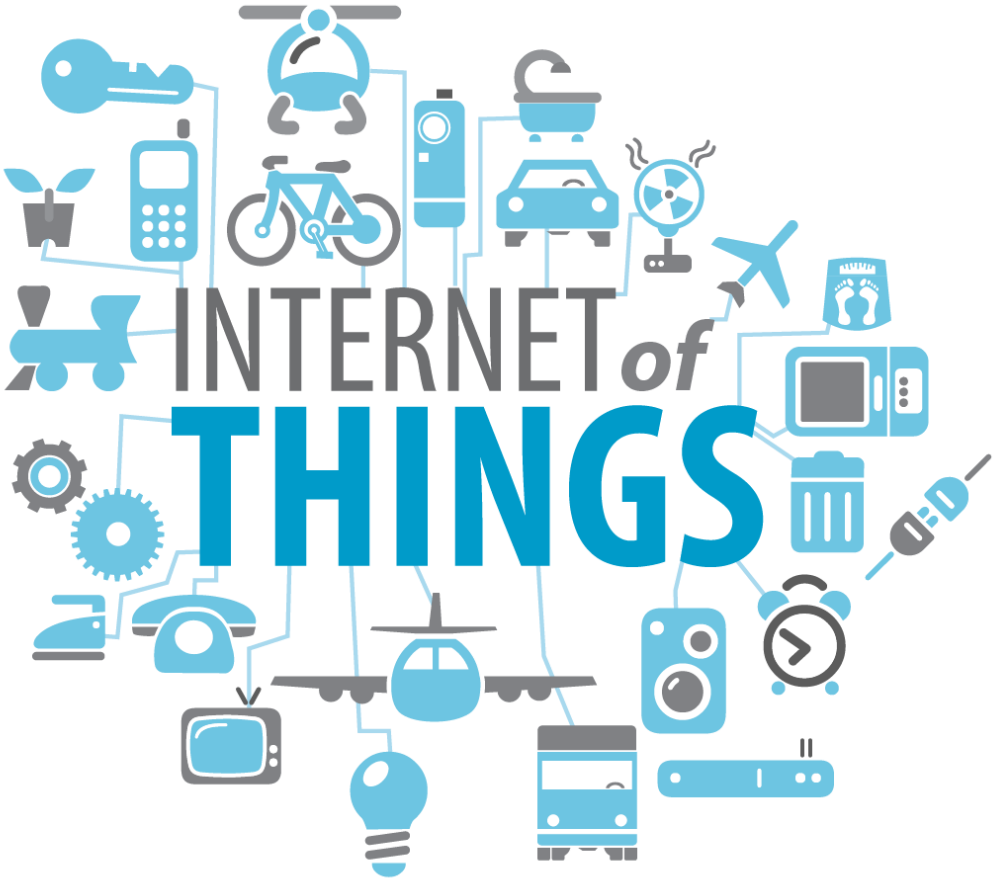
Communication

- transfer of information **from one point** in space-time **to the other**

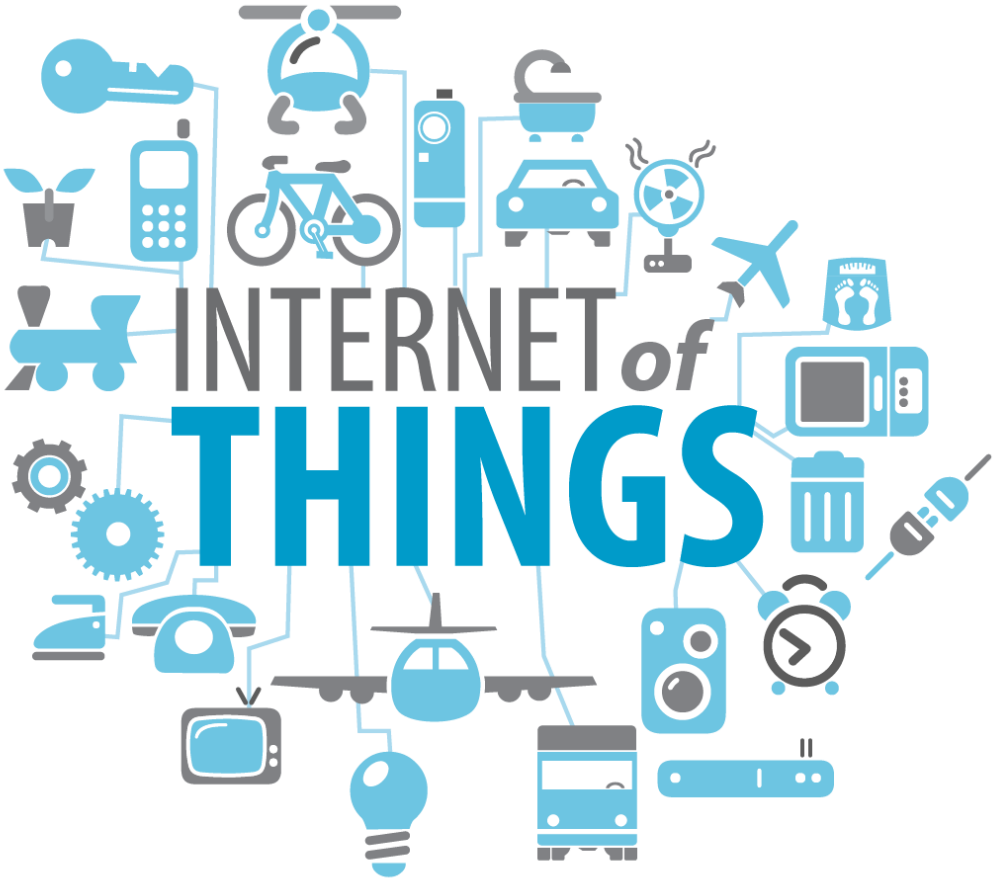
Wireless communication

- the **fundamental limits** of wireless communication are **well understood**

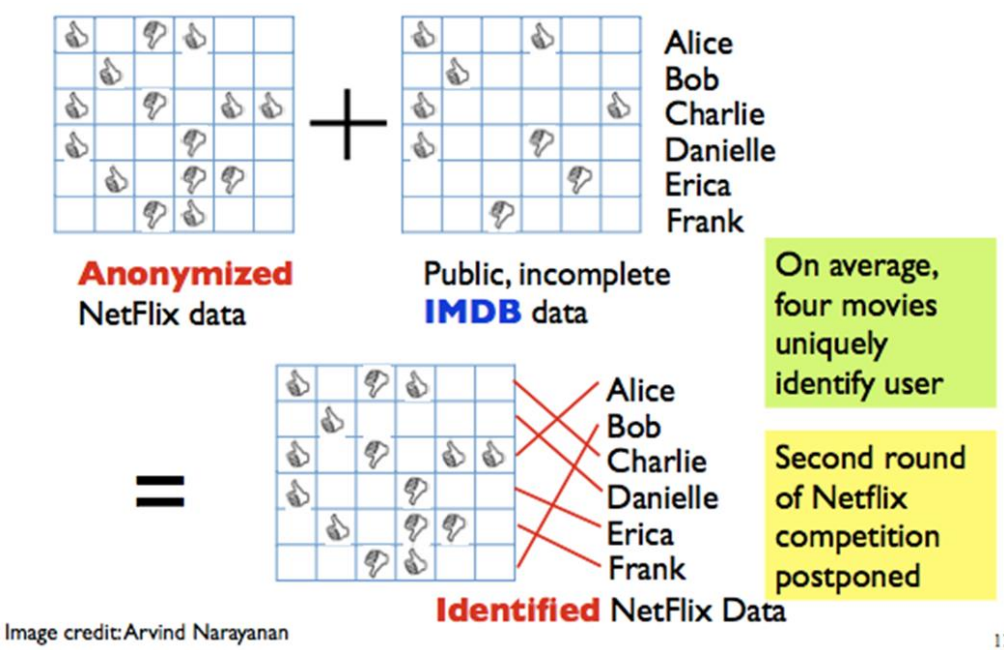
Unprecedented level of connectivity

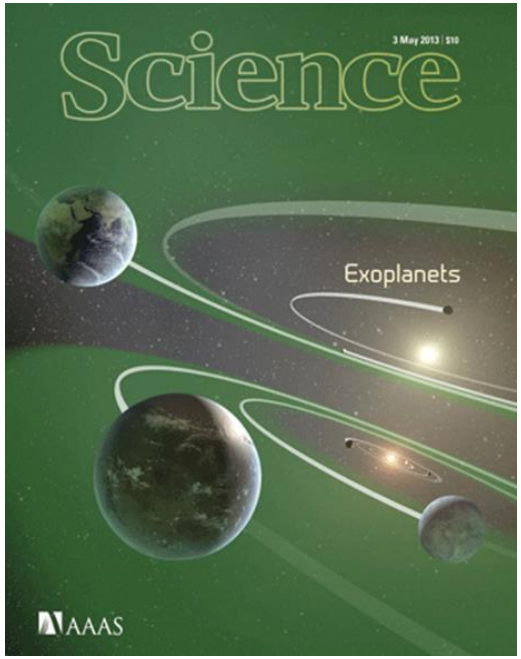


We're being watched!



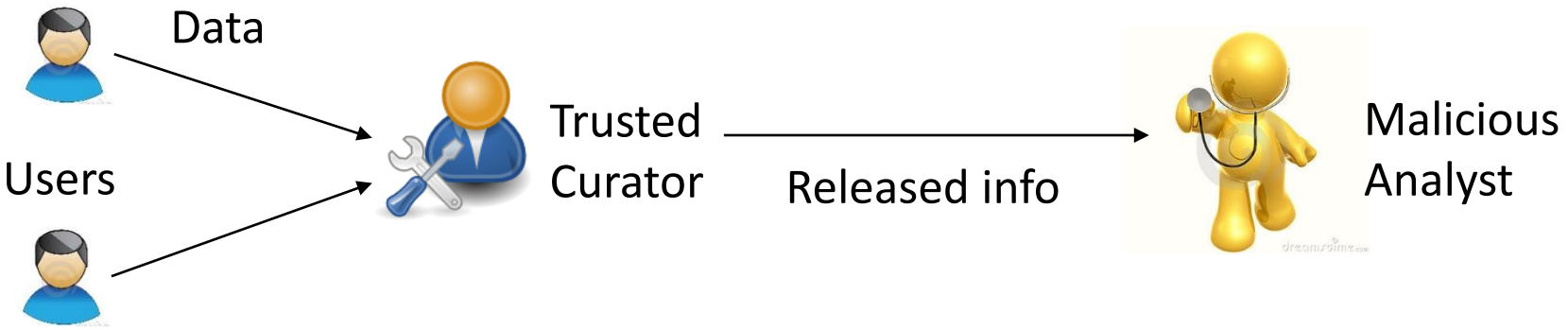
Recent data privacy leaks





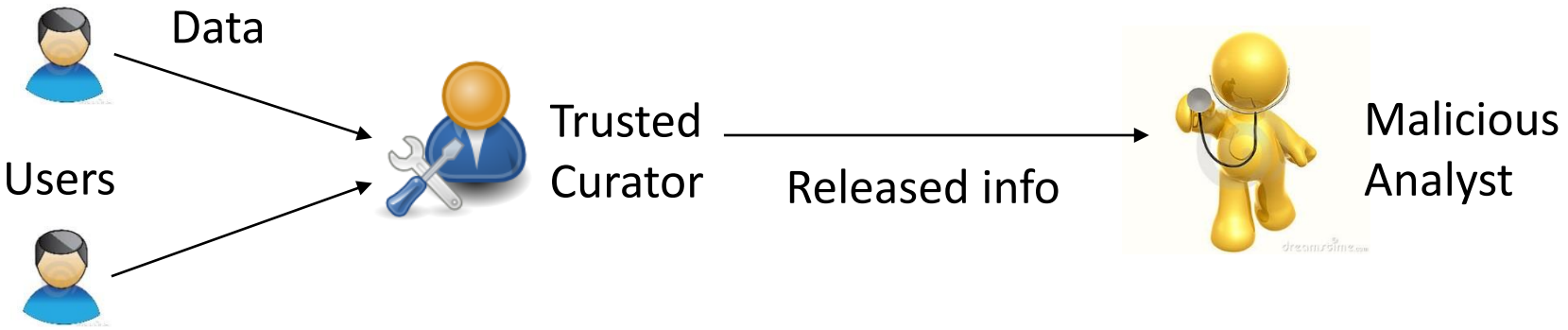
de-anonymizing Netflix data, **identifying** personal genomes

Global privacy model

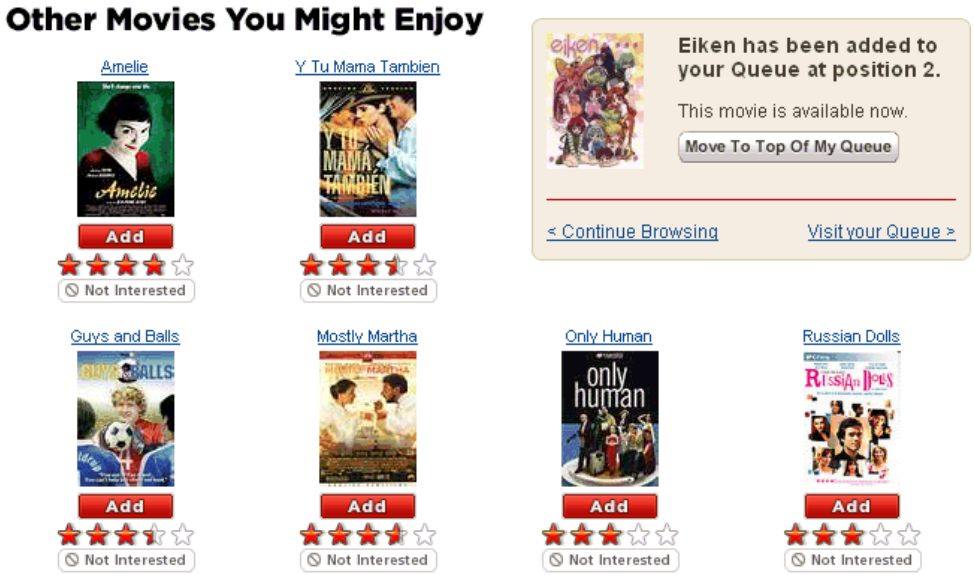


[*Dwork* 2006]

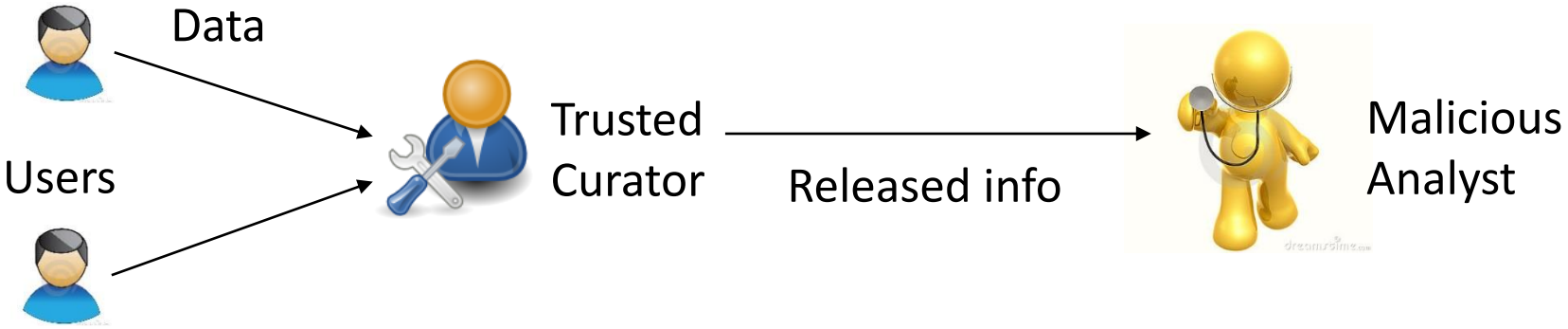
Global privacy model



NETFLIX



Global privacy model



National Institutes of Health

Local privacy model

[*Duchi* 2012]

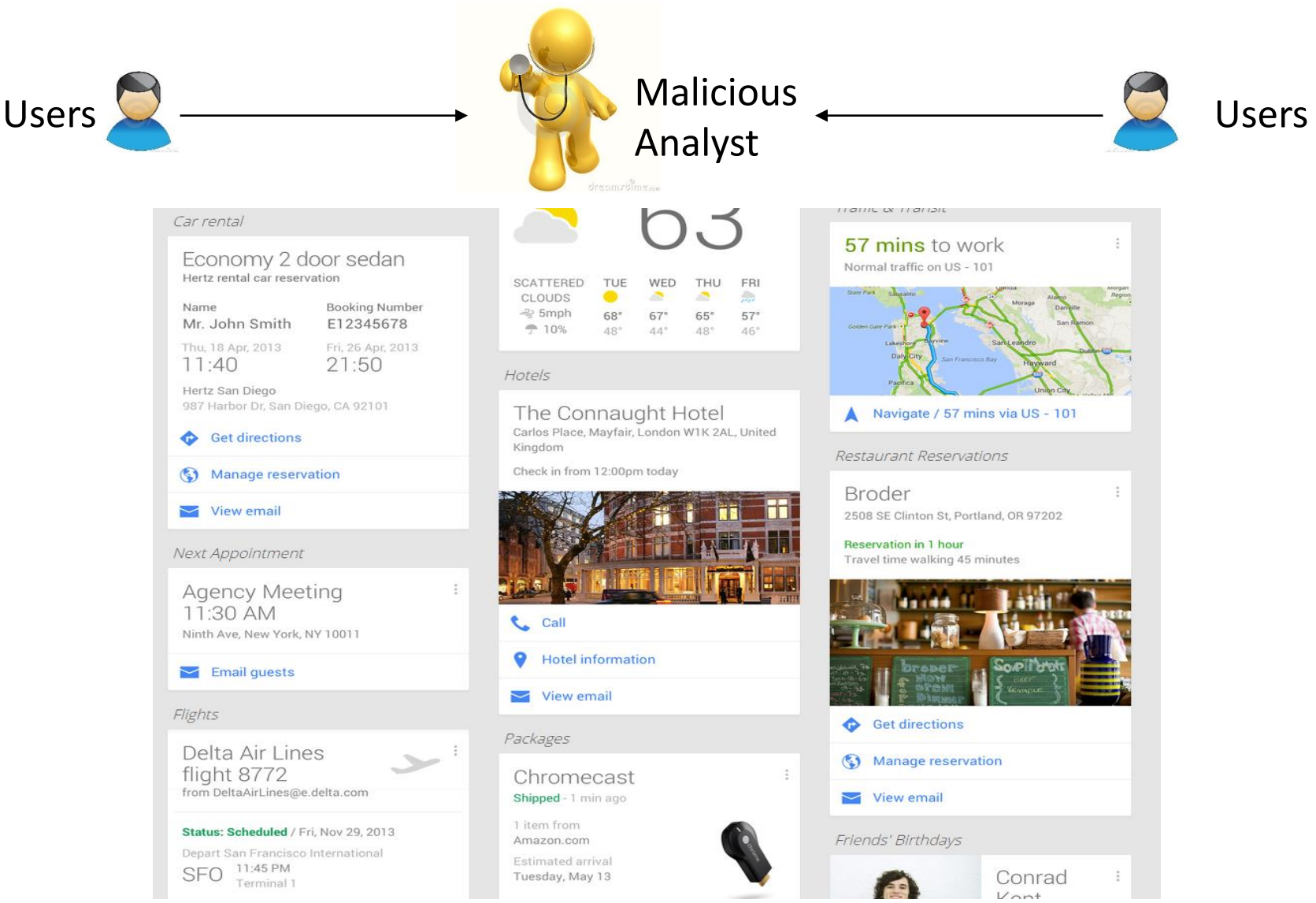
Local privacy model

Google

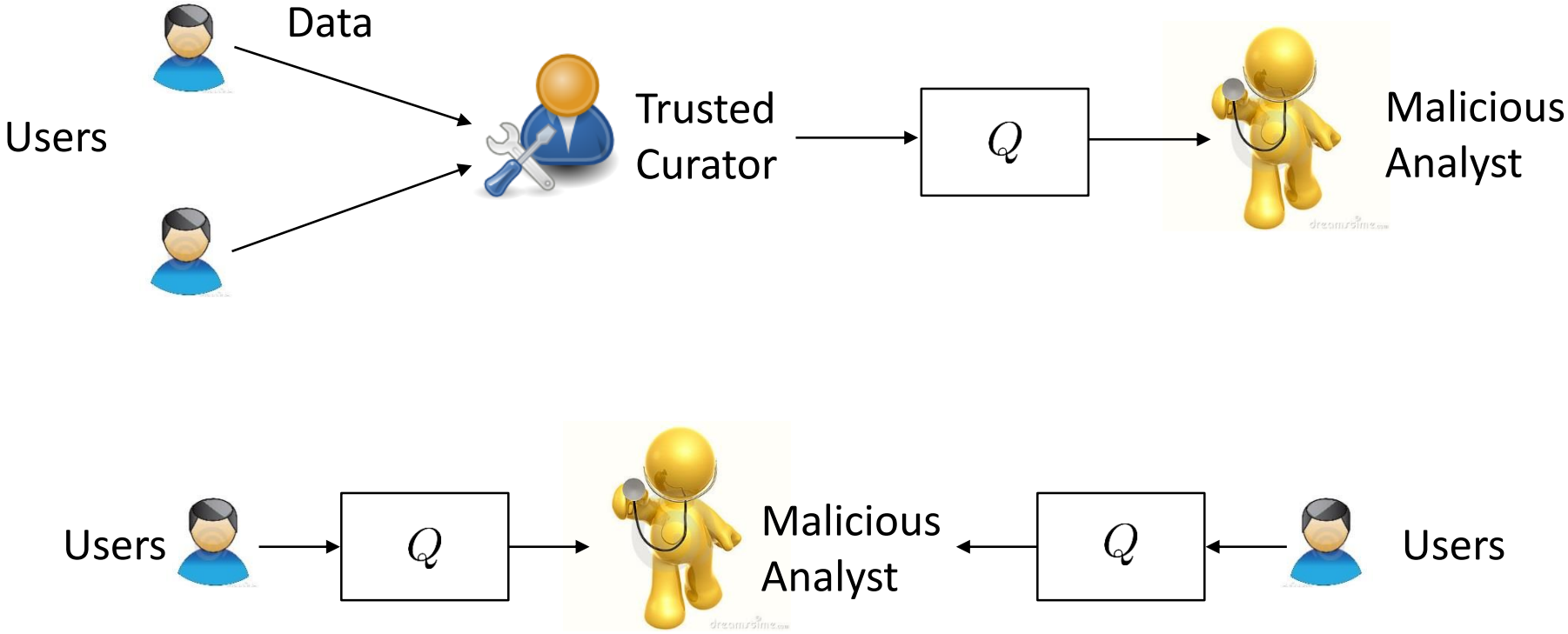
facebook

Microsoft

Local privacy model

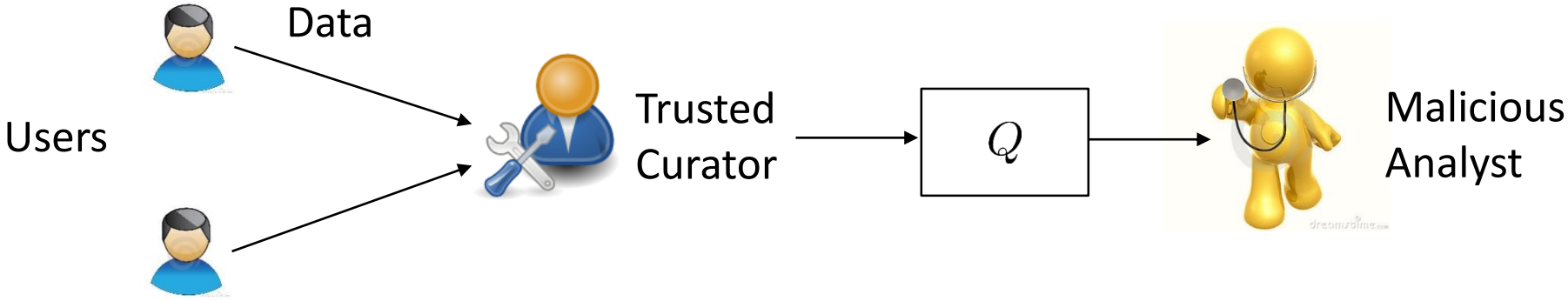


Differential privacy



- Q is a privacy mechanism
- privacy enforced by imposing **differential privacy** parametrized by ε

Differential privacy



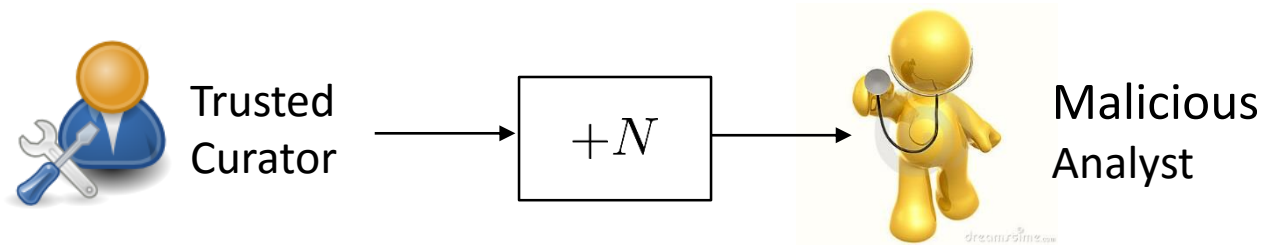
ε controls the level of privacy

large ε, low privacy

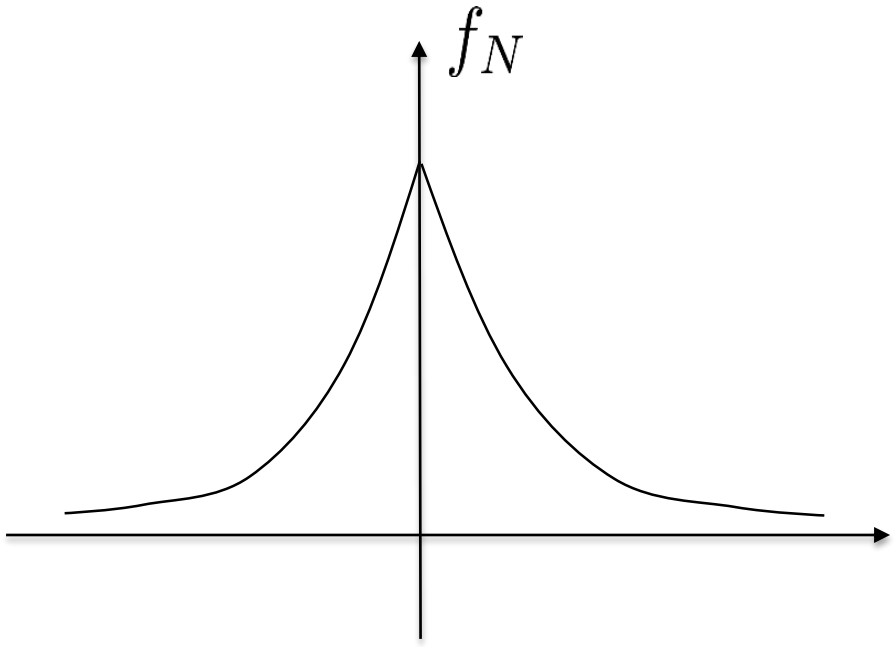
small ε, high privacy

Global Privacy Model

The Laplace mechanism

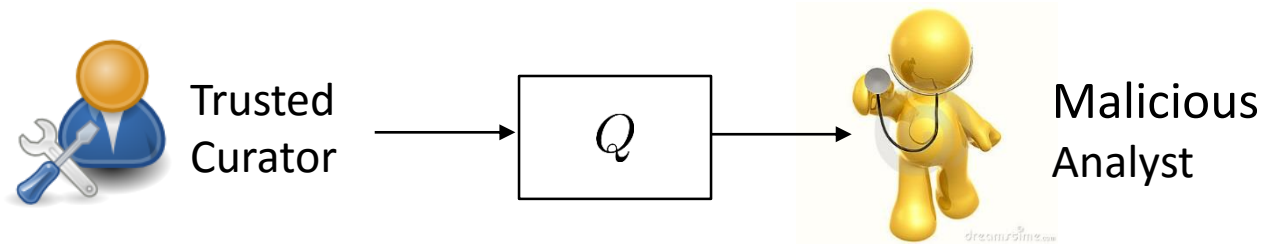


Laplace Mechanism

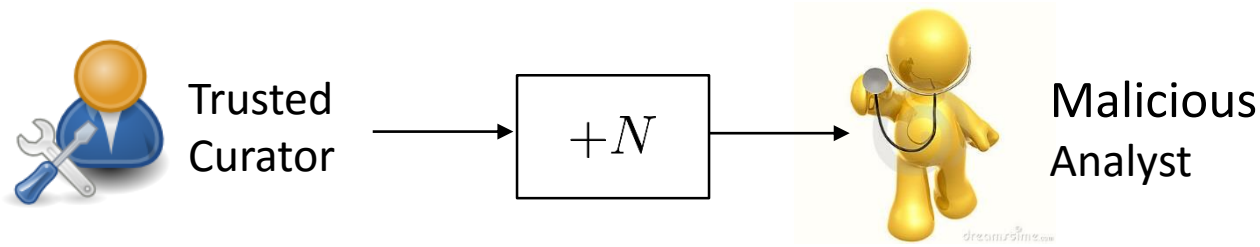


What would Shannon do?

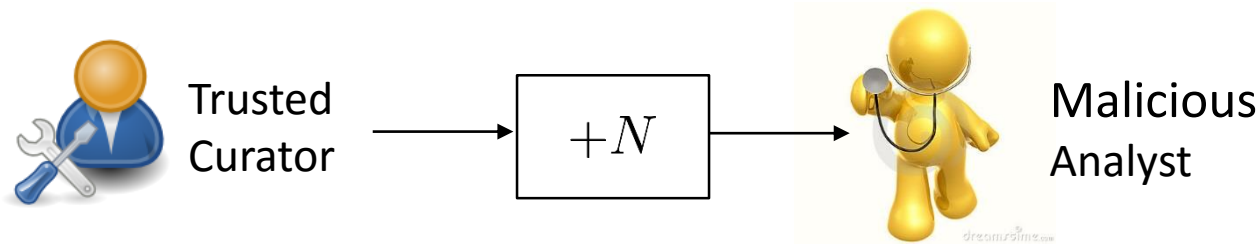
- there is a **fundamental tradeoff** between **privacy** and **utility**



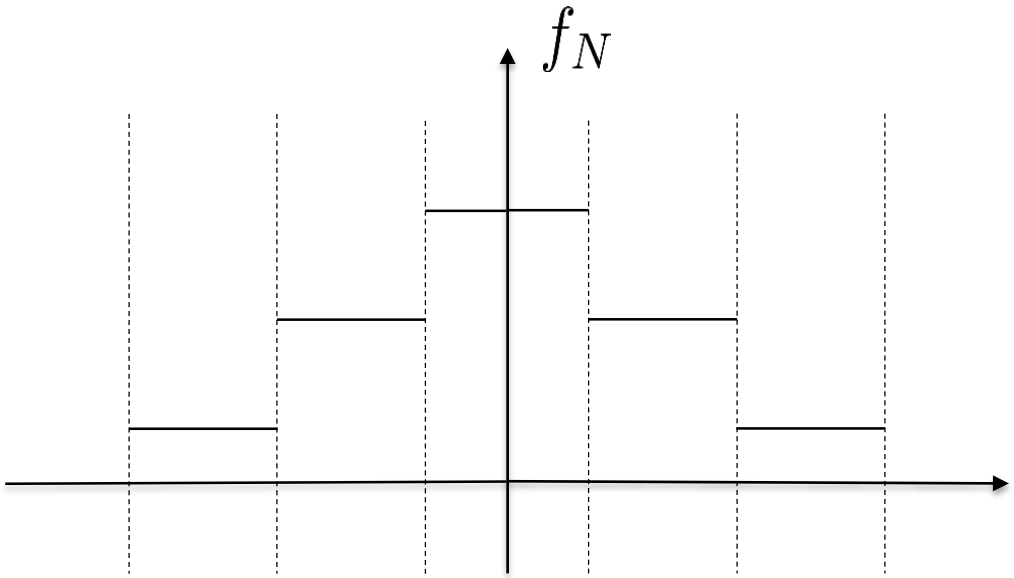
Data independent noise is optimal



Staircase mechanisms are optimal

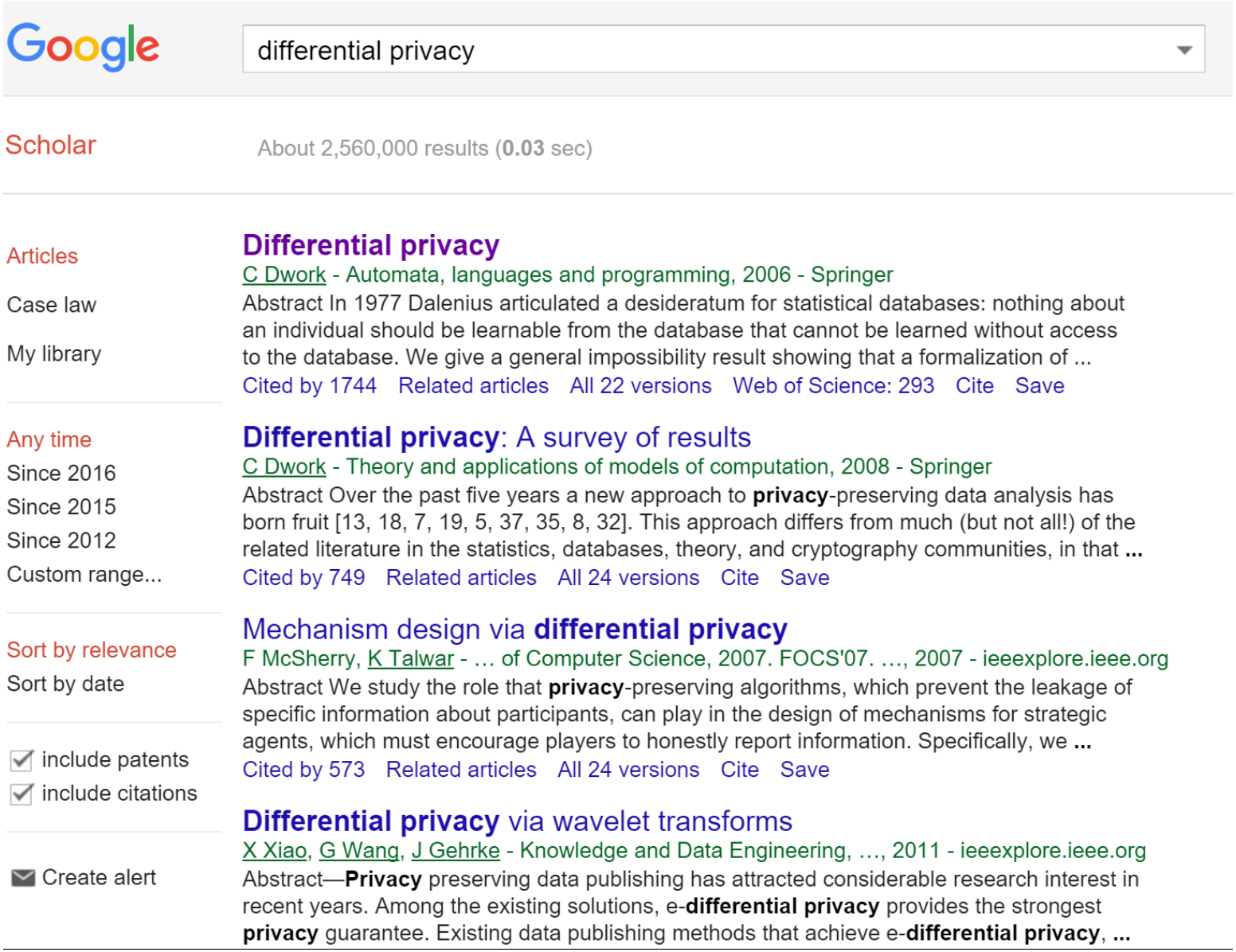


Staircase Mechanism



[TSTSP 15]

Staircase mechanisms are optimal



Local Privacy Model

Local privacy

have you ever used illegal drugs?

Local privacy

have you ever used illegal drugs?

say yes

answer truthfully

[*Warner* 1965]

What would Shannon do?

- there is a **fundamental tradeoff** between **privacy** and **utility**

Main result: binary data

- for binary data:

lie w.p. $\frac{1}{e^{\varepsilon}+1}$

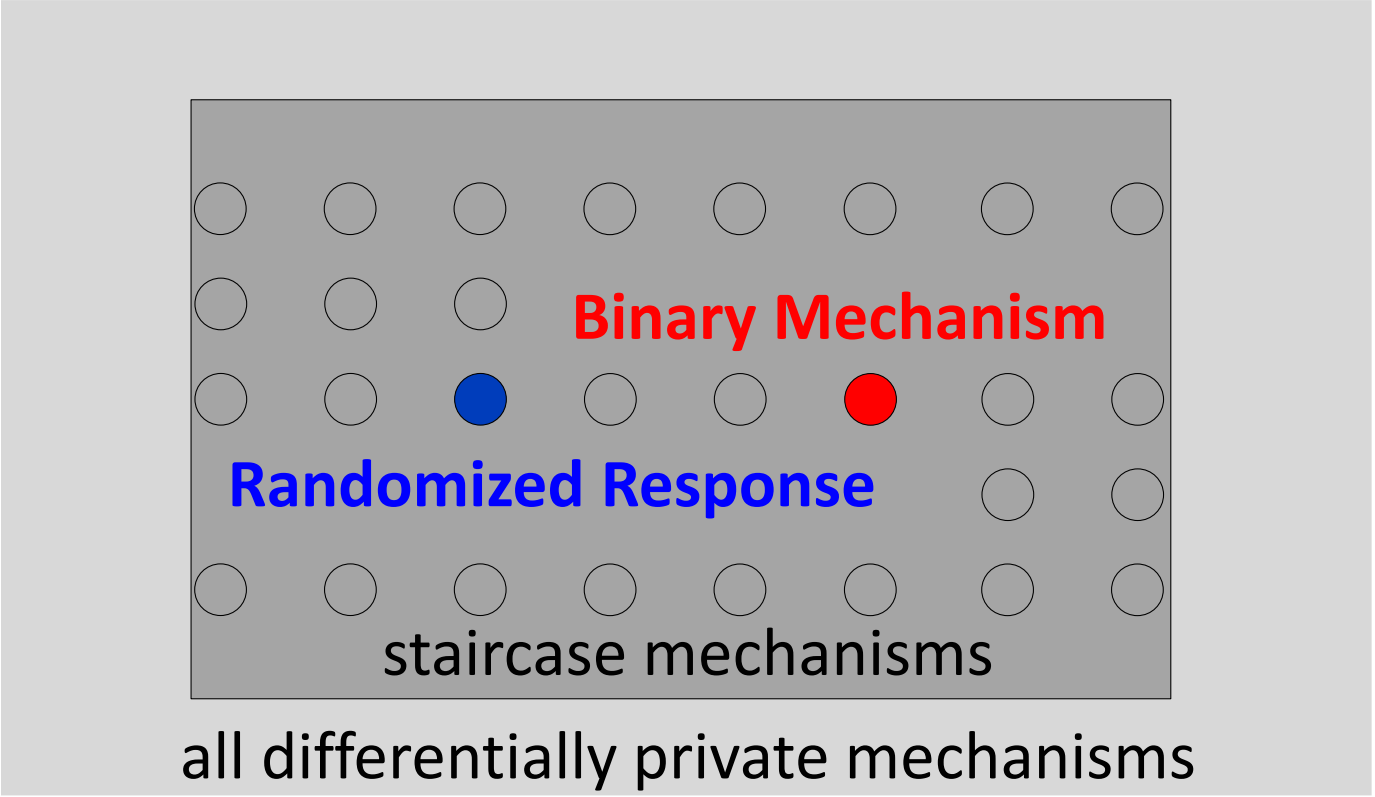
say the truth w.p. $\frac{e^{\varepsilon}}{e^{\varepsilon}+1}$

- optimal for **all utilities obeying the data processing inequality**

[*JMLR 2015*, NIPS 14]

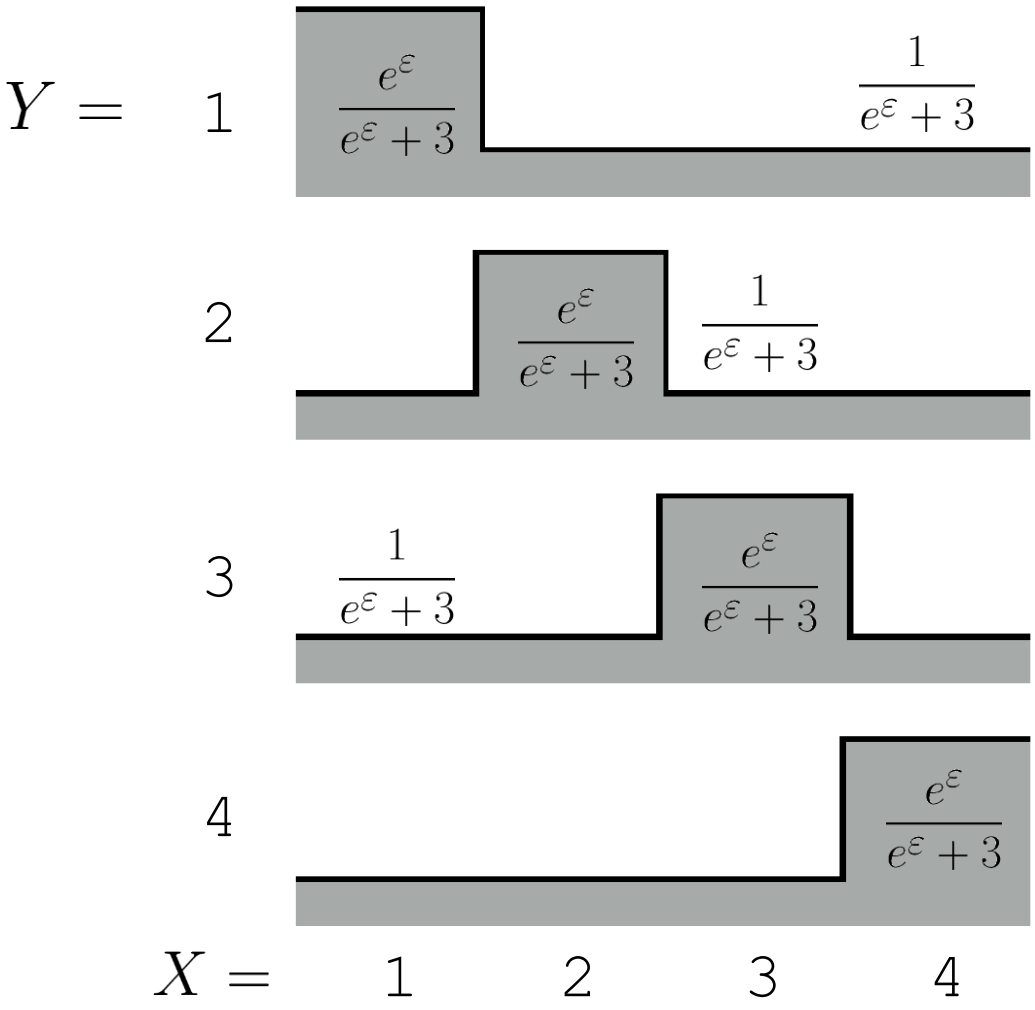
Main result: general data

- for general k-ary data:



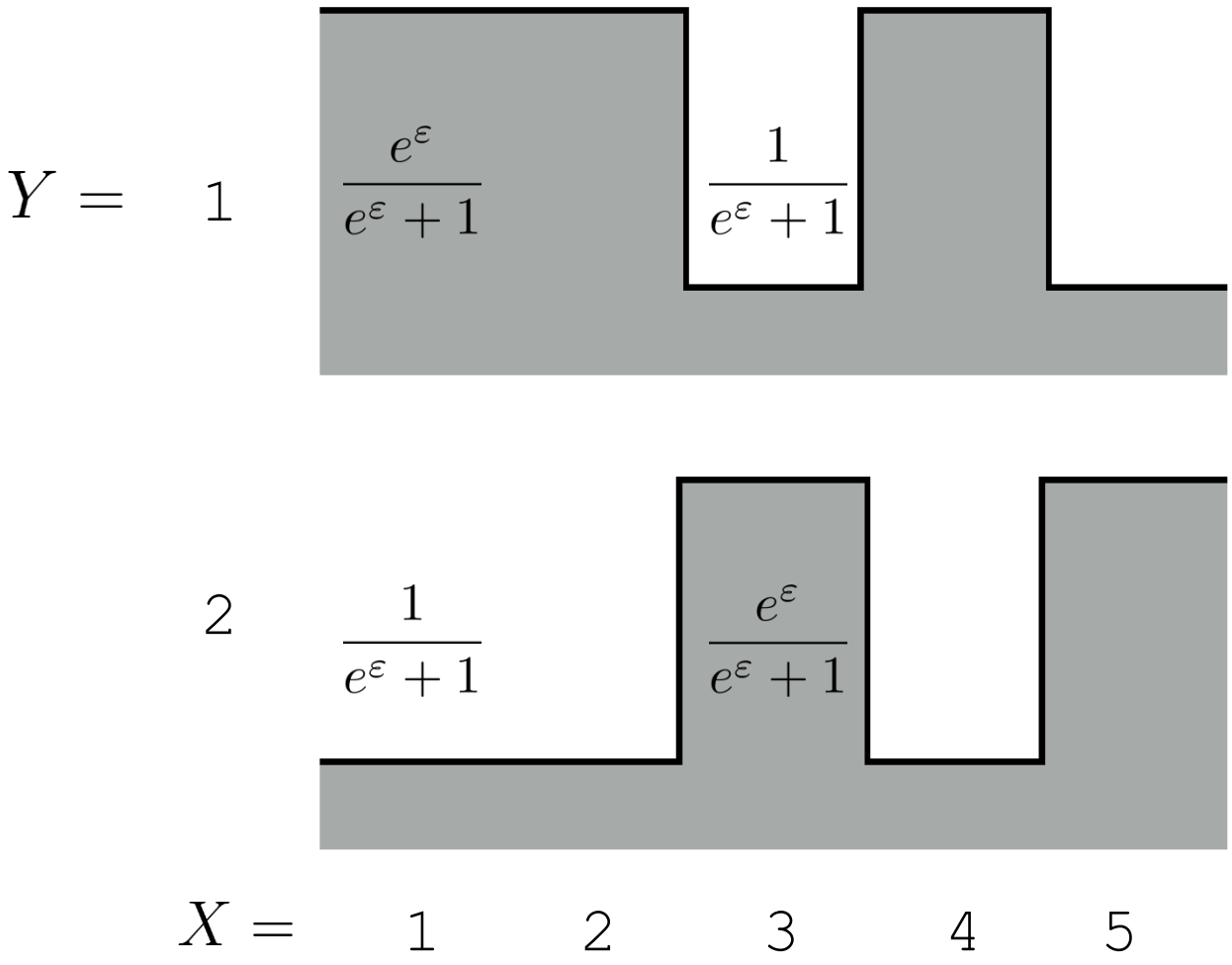
[NIPS 14, *JMLR 15*]

Randomized Response



- optimal in the **low privacy regime**

Binary mechanism

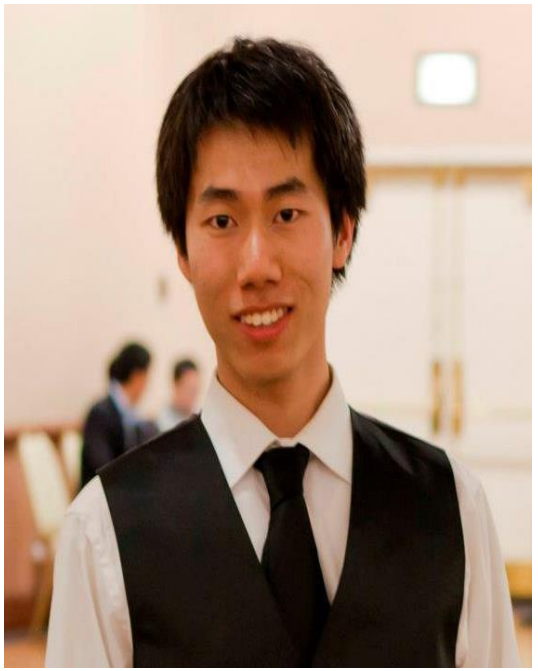


- optimal in the **high privacy regime**

Acknowledgments

Sewoong Oh

Pramod Viswanath



Quan Geng