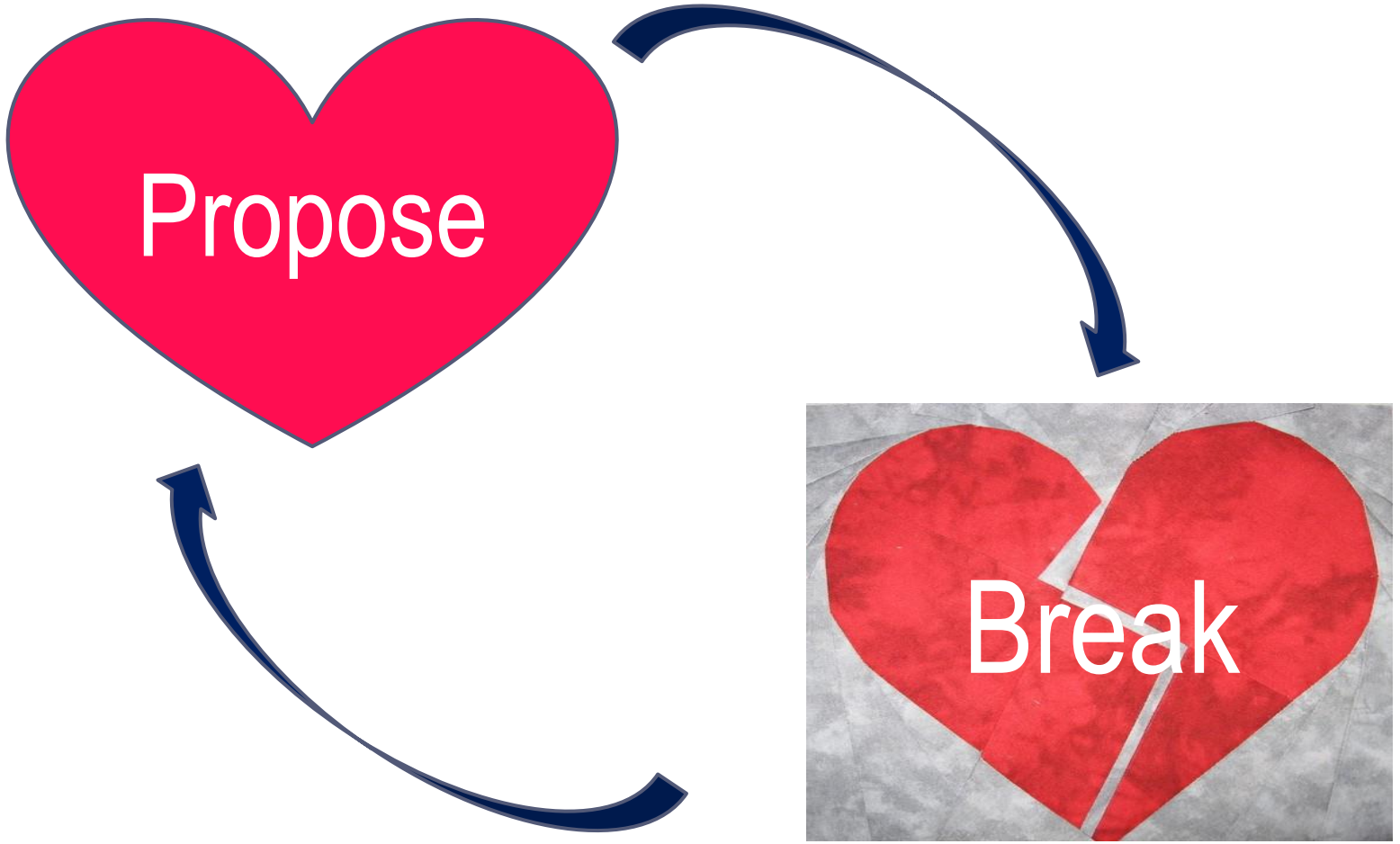
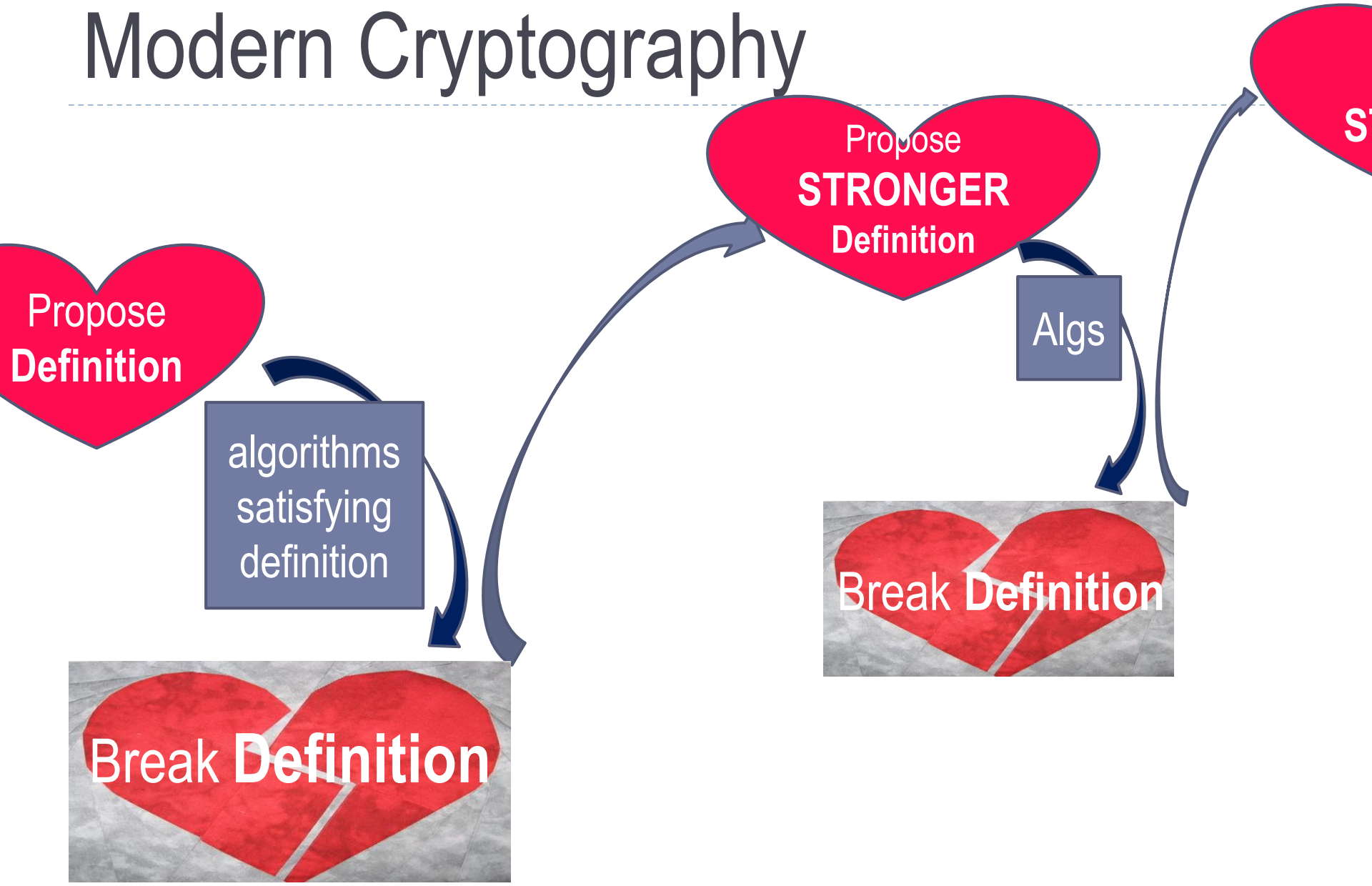
Discussant: The Secret Life of I.J. Good

Cynthia Dwork, Harvard University

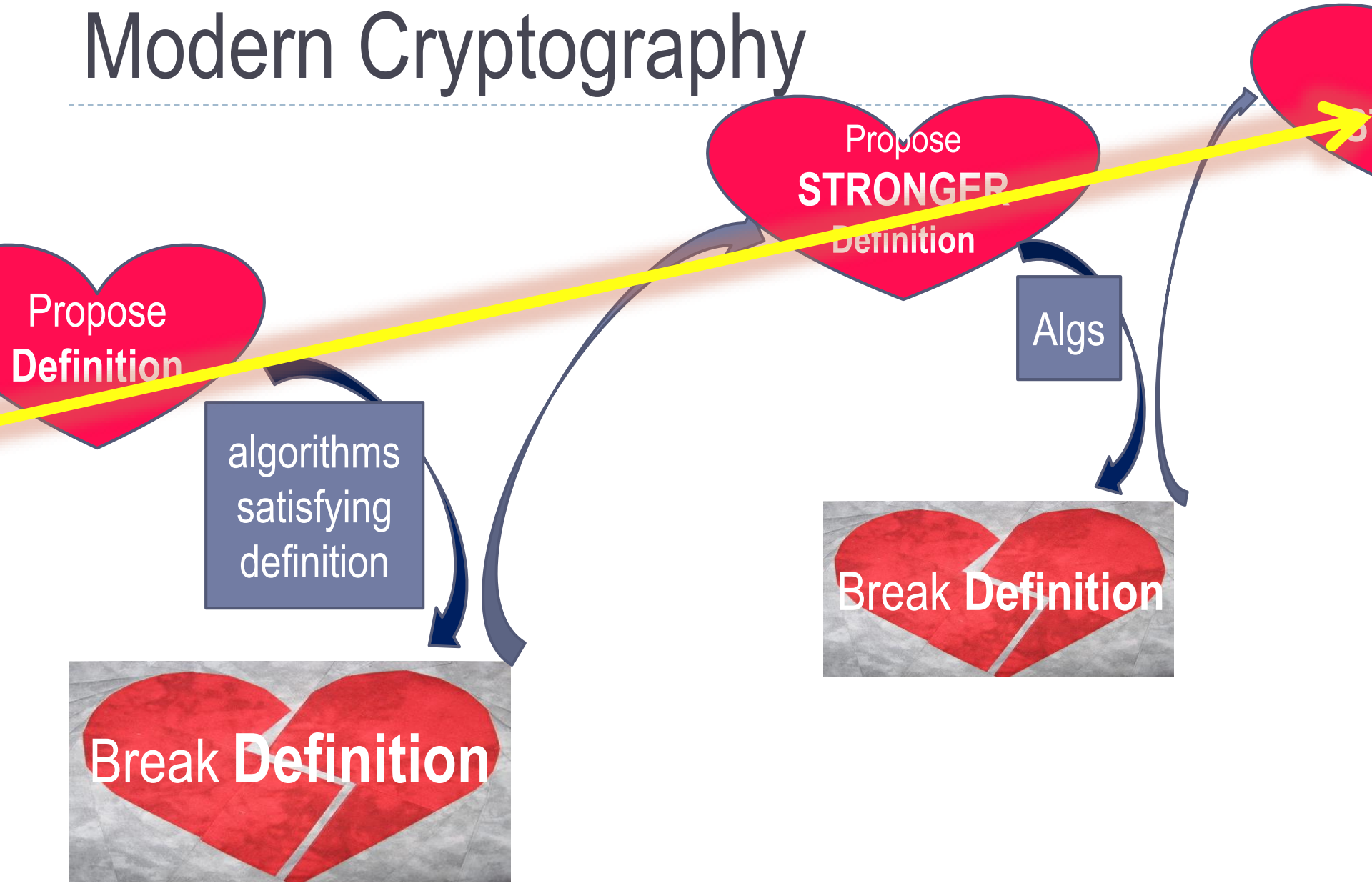
Pre-Modern Cryptography



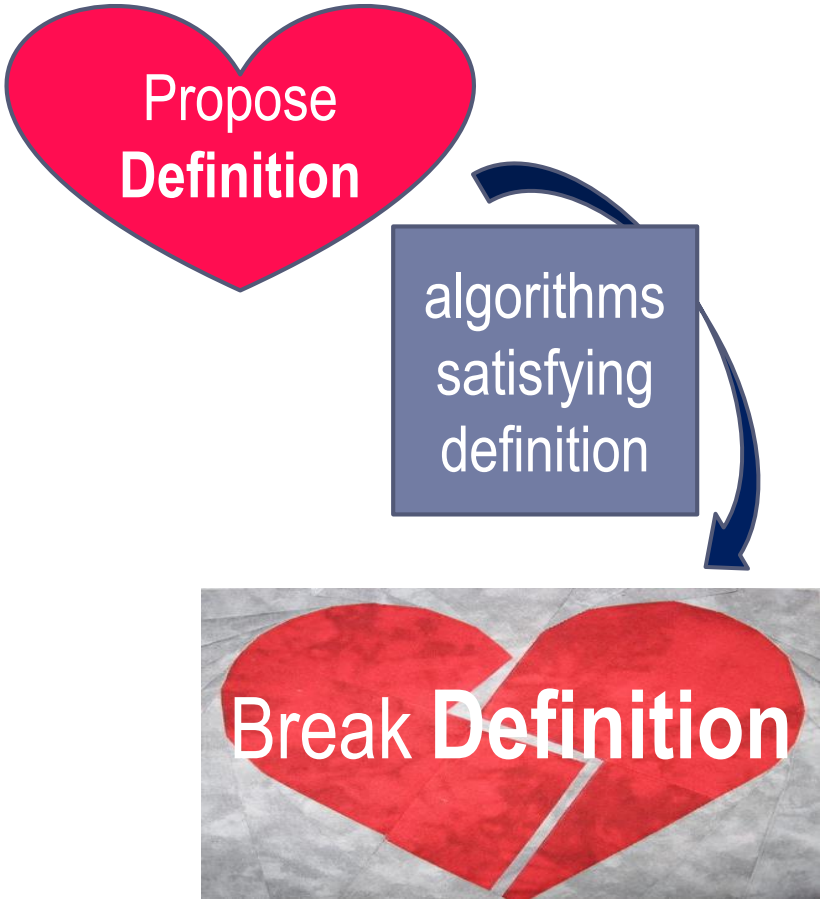
Modern Cryptography



Modern Cryptography



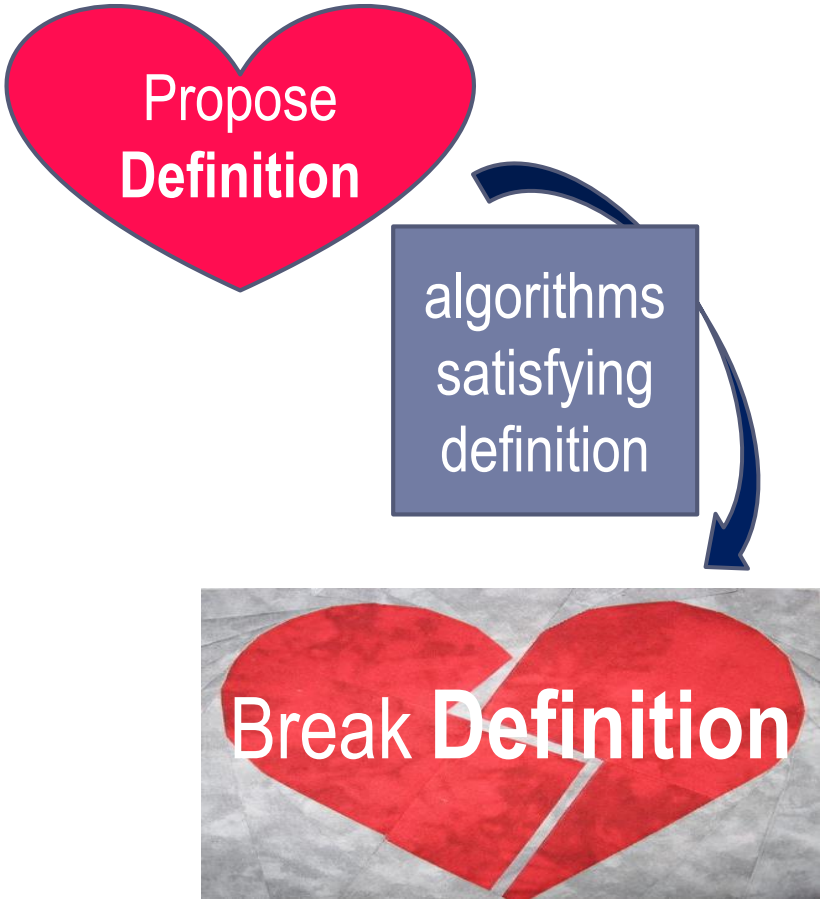
Security for an Encryption Scheme



- What is the goal of the adversary?
- What are the resources available to the adversary?

Goldwasser and Micali, 1982

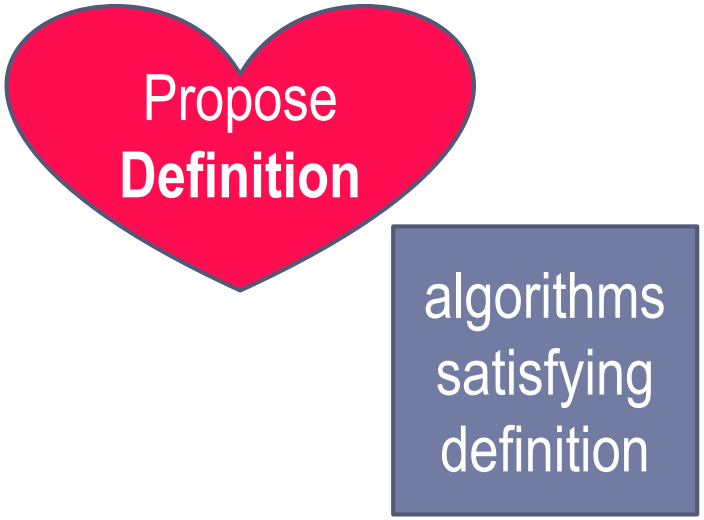
Resources



- What is the goal of the adversary?
- What are the resources available to the adversary?
  - Adversary runs in polynomial time (in a security parameter κ)
  - Can have arbitrary auxiliary information
  - For today: adversary is a passive eavesdropper (ie, no chosen ciphertext attacks)

Goldwasser and Micali, 1982

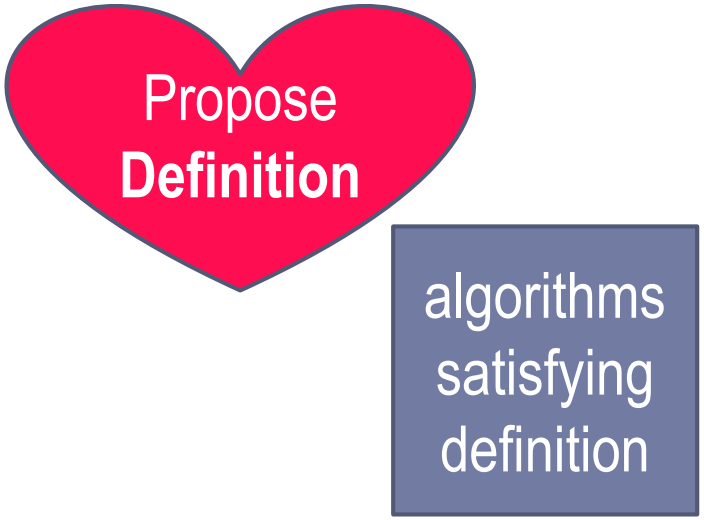
Goal 1: Distinguishing from Random



- What is the goal of the adversary?
  - Adversary choses message m. Attempts to distinguish encryptions of m from encryptions of random strings of the same length.

Goldwasser and Micali, 1982

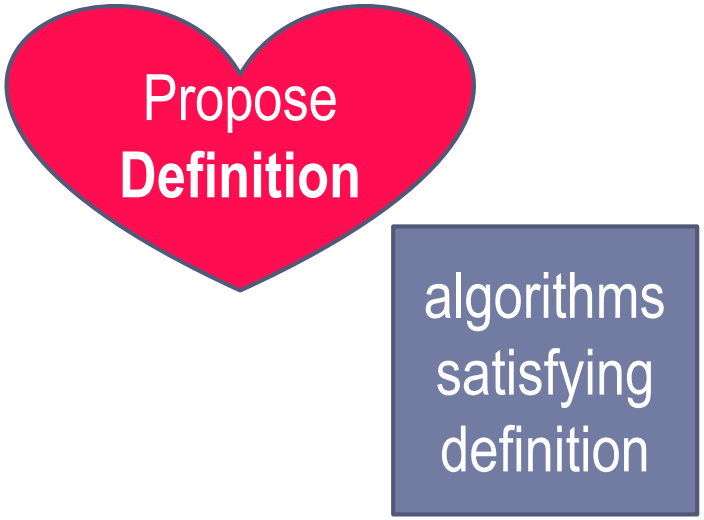
Goal 2: Distinguishing Two Messages



- What is the goal of the adversary?
  - Adversary choses messages m_0, m_1 of the same length. Attempts to distinguish encryptions of m_0 from encryptions of m_1.

Goldwasser and Micali, 1982

Goal 3: Semantic Security



- What is the goal of the adversary?
  - Adversary choses a distribution D on messages and function f.
  - Adversary tries to guess $f(m)$, where $m \sim D$
  - … better than it could do without access to the ciphertext
- All three goals are equivalent

Goldwasser and Micali, 1982

Formalizing Semantic Security

- $\{(KeyGen_\kappa, Enc_\kappa, Dec_\kappa)\}_{\kappa=1,\ldots}$
- Everything runs in time polynomial in κ; all lengths are poly(κ)
  - $KeyGen_\kappa$ produces key(s). If public-key system, A receives E.
  - $A(E)$ chooses a distribution D on messages and arbitrary function f
  - $m \sim D$ is chosen and the ciphertext α is presented to A.
  - A is given length of m and arbitrary auxiliary information: $z = (z(m), |m|)$.
  - $A(E, \alpha, z)$ outputs guess of $f(m)$.

- $\forall A \exists S$ such that $\forall$ poly p:

$$|\Pr[A(E, \alpha, z) \text{ guesses } f(m)] - \Pr[S(E, z) \text{ guesses } f(m)]| \leq 1/p(\kappa)$$

Probabilities taken over everything

Goldwasser and Micali, 1982

Formalizing Semantic Security

- $\{(KeyGen_\kappa, Enc_\kappa, Dec_\kappa)\}_{\kappa=1,\ldots}$
- Everything runs in time polynomial in κ; all lengths are poly(κ)
  - $KeyGen_\kappa$ produces key(s). If public-key system, A receives E.
  - $A(E)$ chooses a distribution D on messages and arbitrary function f
  - $m \sim D$ is chosen and the ciphertext α is presented to A.
  - A is given length of m and arbitrary auxiliary information: $z = (z(m), |m|)$.
  - $A(E, \alpha, z)$ outputs guess of $f(m)$.

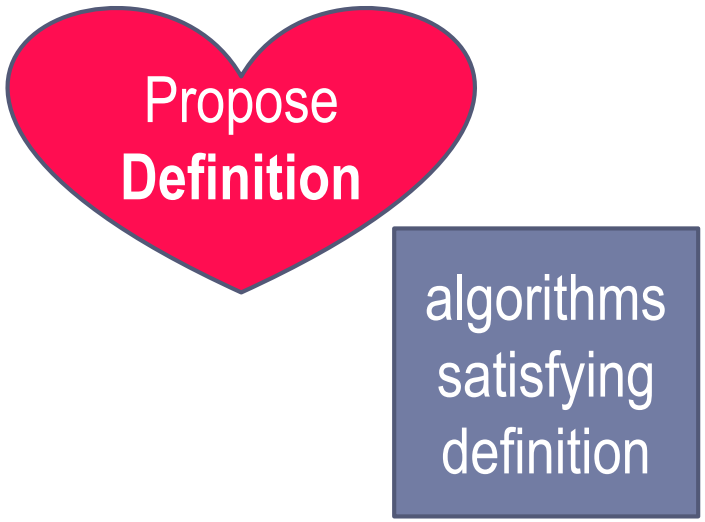
- $\forall A \exists S$ such that $\forall$ poly p:

$$|\Pr[A(E, \alpha, z) \text{ guesses } f(m)] - \Pr[S(E, z) \text{ guesses } f(m)]| \leq 1/p(\kappa)$$

"The ciphertext effectively leaks no information about m"

Goldwasser and Micali, 1982

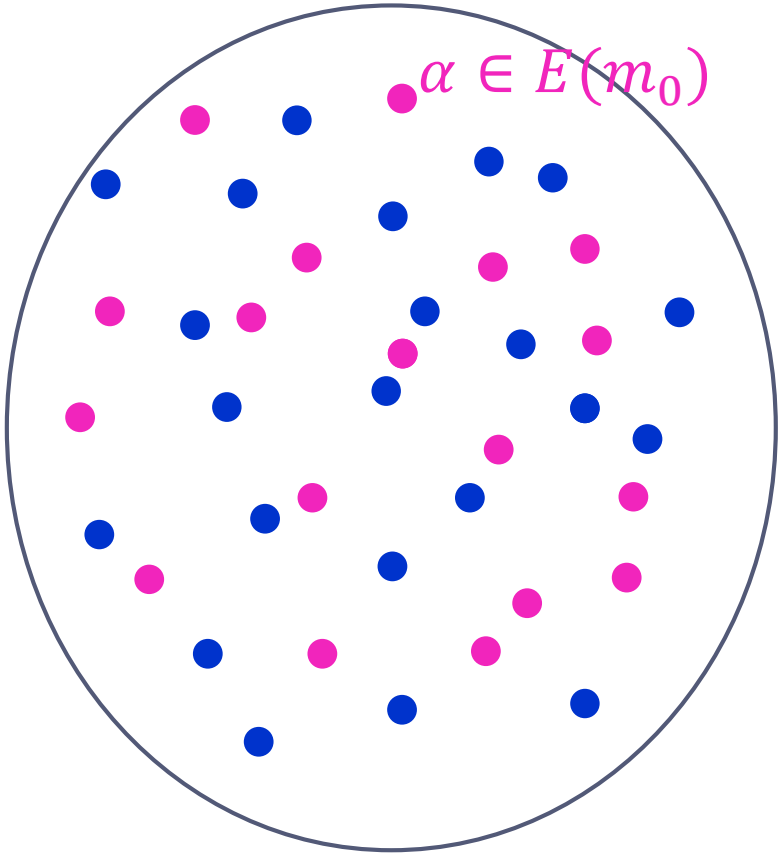
Algorithms



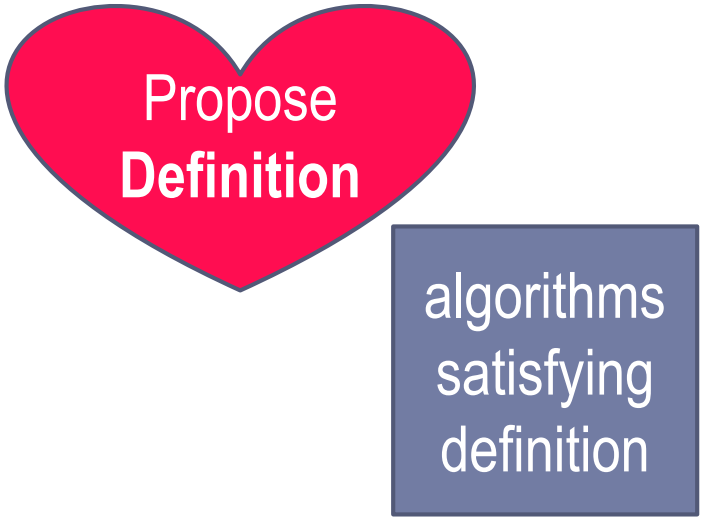
- Base an algorithm on a hard* problem
  - Prove that any efficient adversary achieving its goal can be converted into an efficient algorithm to solve the hard problem (showing that the problem is not hard after all)
- Examples: quadratic residuousity, factoring, finding discrete logarithms, finding a short basis for a lattice, learning with errors

Infinite Bayes Factor

- $\frac{\Pr[\alpha|m_0]}{\Pr[\alpha|m_1]} = \infty$
- Can enumerate all blue
- Can enumerate all pink
- Without knowledge of the decryption key: Cannot efficiently determine color of point without efficiently solving the underlying mathematical problem

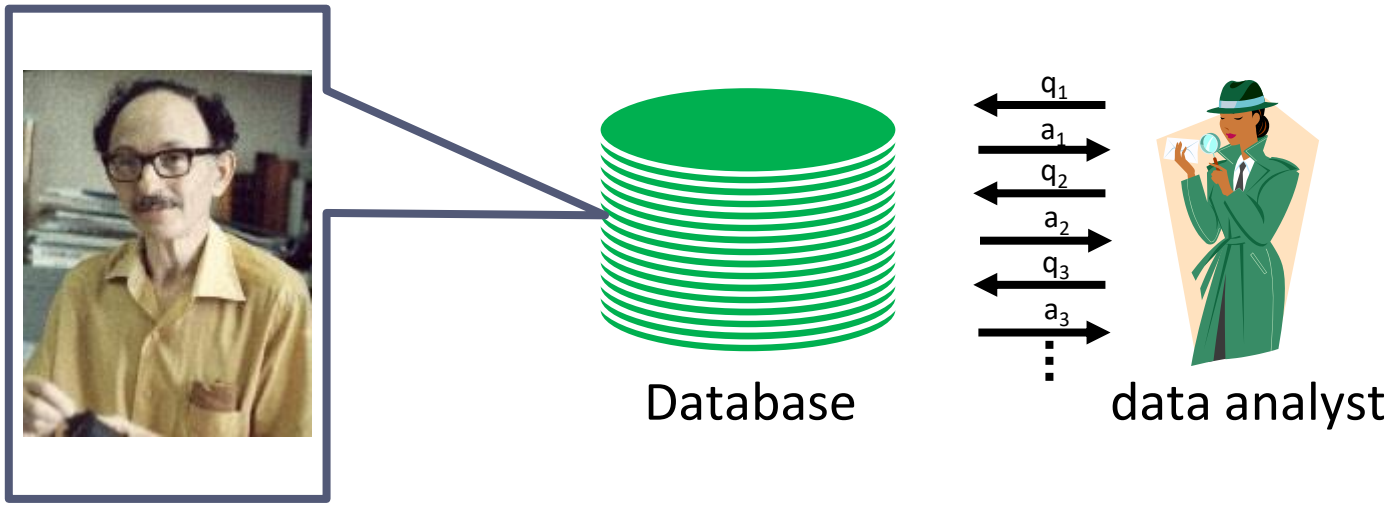


Randomness

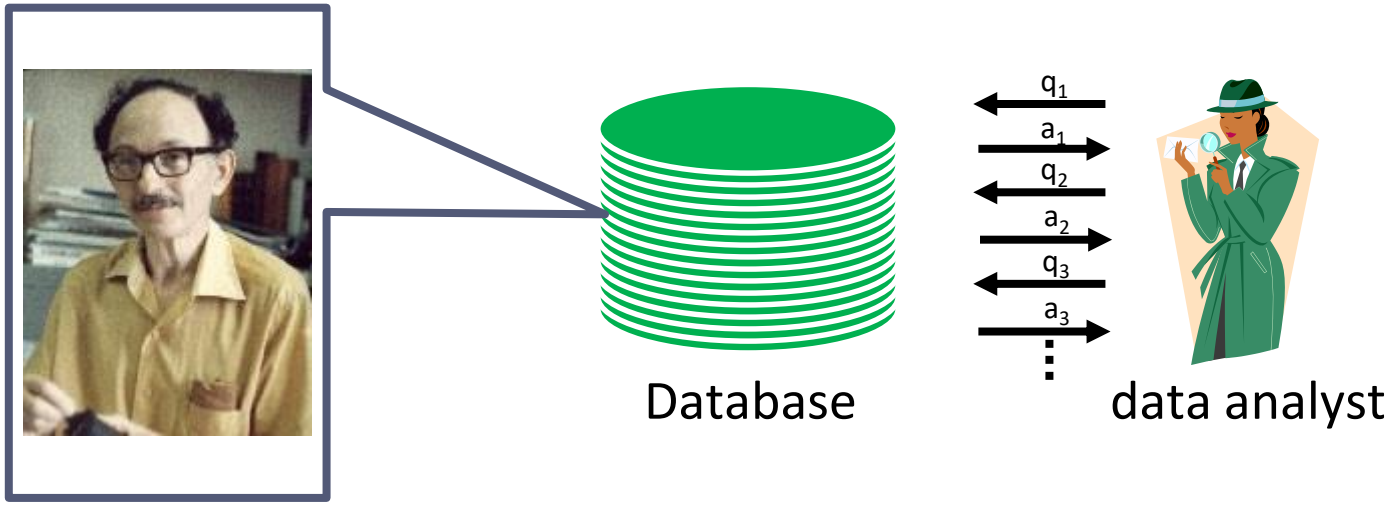


- A few high-quality random bits suffice
- Pseudo-randomness can be based on any one-way function

Privacy-Preserving Data Analysis

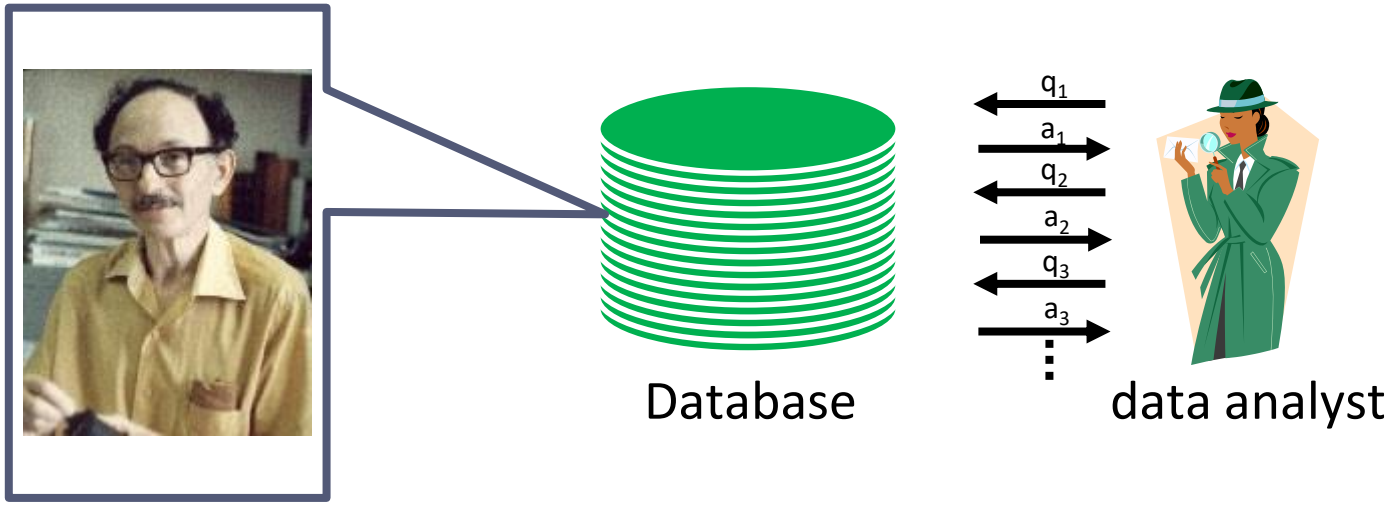


Privacy-Preserving Data Analysis?



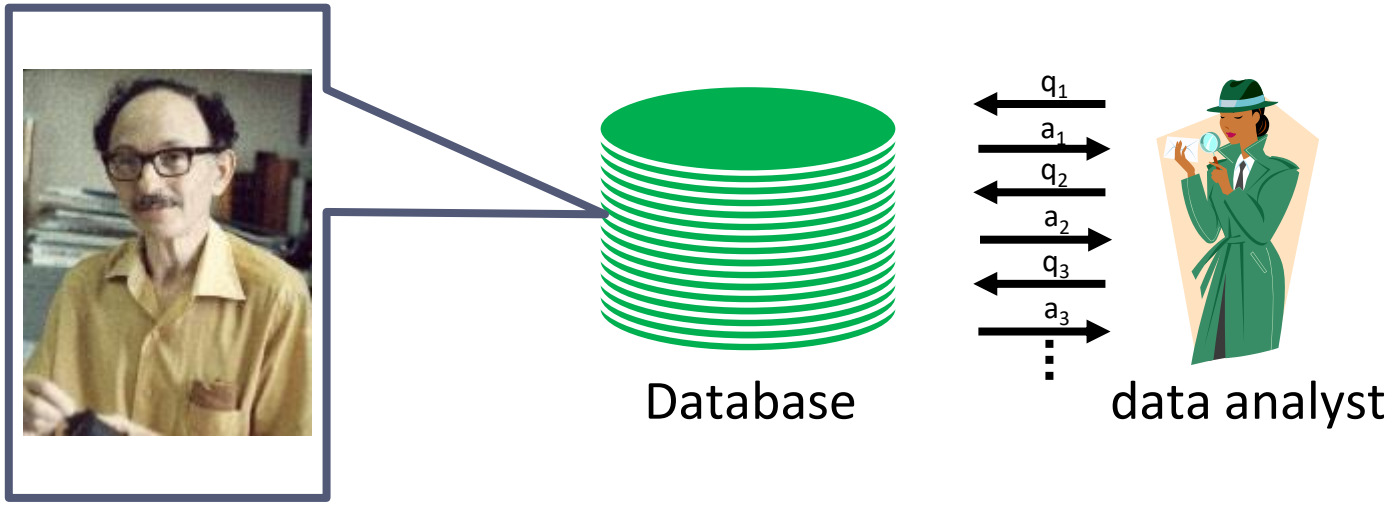
- "Can't learn anything new about Good"?
- Dalenius, 1977; Goldwasser and Micali: semantic security 1982

Privacy-Preserving Data Analysis?



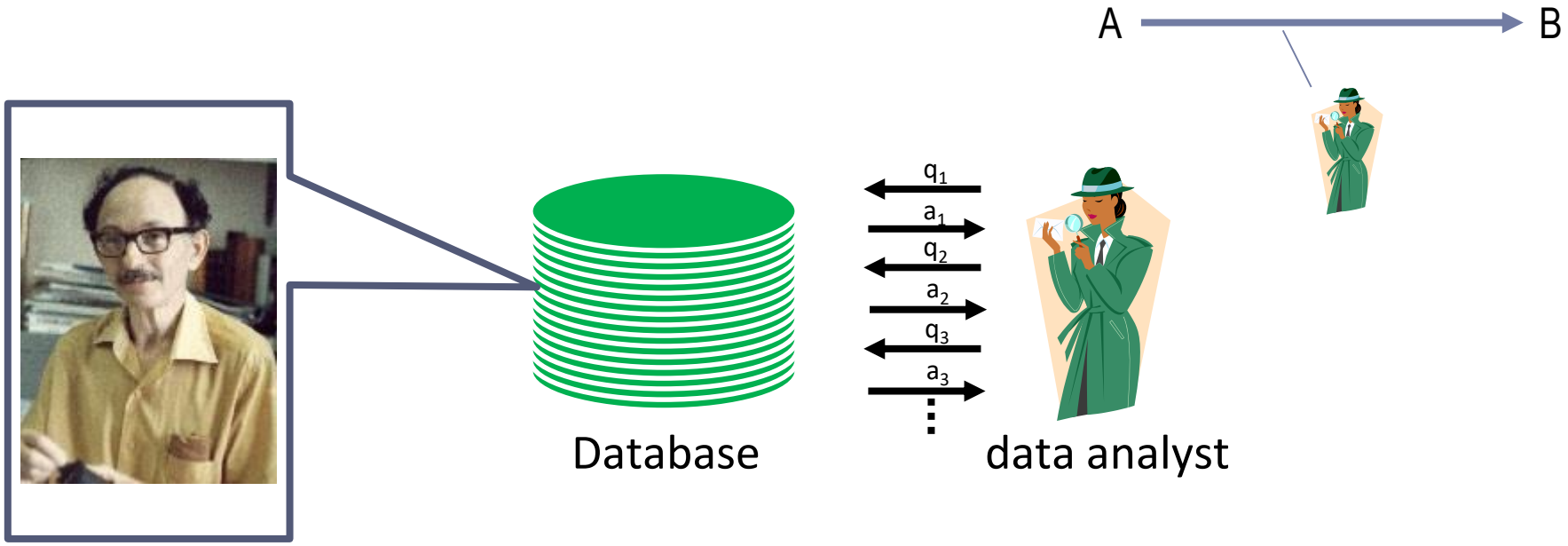
- "Can't learn anything new about Good"?
- Then what is the point?

Privacy-Preserving Data Analysis?



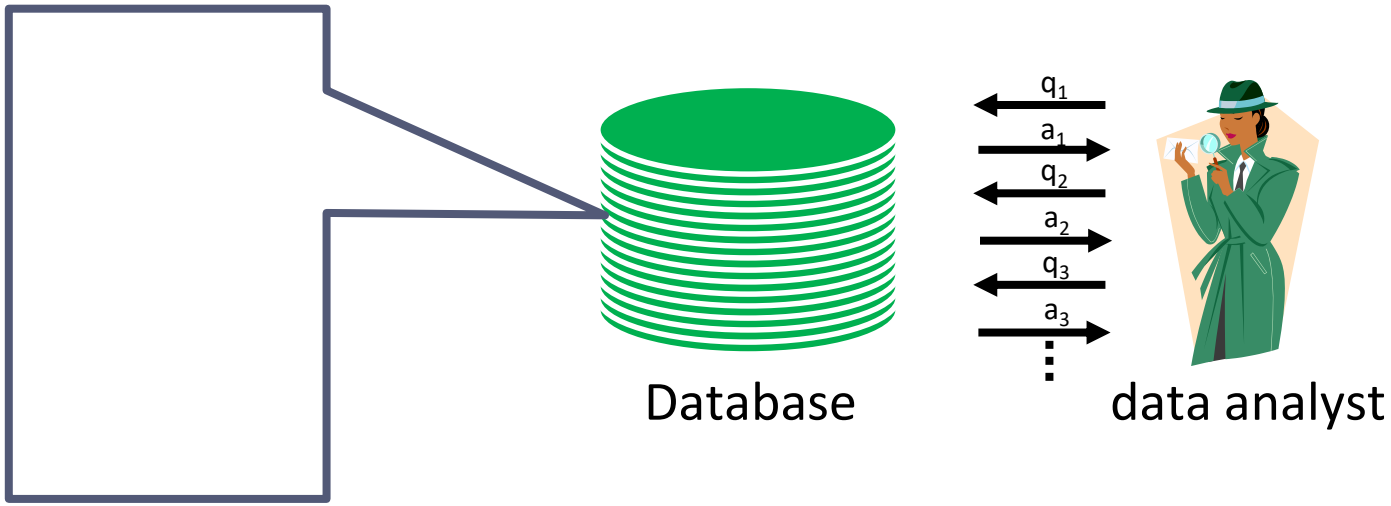
- "Can't learn anything new about Good"?
- Then what is the point?

Privacy-Preserving Data Analysis?



- Semantic Security is the Wrong Notion!

Key Observation



- Learn the same things even without Good

Differential Privacy

- The outcome of any analysis is essentially equally likely, independent of whether any individual joins, or refrains from joining, the dataset.

Differential Privacy

M gives ϵ-differential privacy if for all pairs of adjacent data sets x, y, and all events S

$$\Pr[M(x) \in S] \leq e^{\epsilon} \Pr[M(y) \in S]$$

Randomness introduced by M

Dwork, McSherry, Nissim, and Smith, 2006

Differential Privacy Bounds the Bayes Factor

M gives ϵ-differential privacy if for all pairs of adjacent data sets x, y, and all events S

$$\Pr[M(x) \in S] \leq e^{\epsilon} \Pr[M(y) \in S]$$

$$\Pr[S|x] \leq e^{\epsilon} \Pr[S|y]$$

$$\frac{\Pr[S|x]}{\Pr[S|y]} \leq e^{\epsilon}$$

Dwork, McSherry, Nissim, and Smith, 2006

Privacy for Databases that Teach

- Database teaches that smoking causes cancer.
  - Smoker S's insurance premiums rise.
  - Premiums rise even if S not in database!
- Learning that smoking causes cancer is the whole point.
  - Smoker S enrolls in a smoking cessation program.

- Differential privacy: limit harms to the teachings, not participation

Dwork, McSherry, Nissim, and Smith, 2006

Bayes Factors Large and Small

- Modern Cryptography
  - Complexity theory provides an end run around infinite Bayes factor
- Differentially Private Data Analysis
  - Controls the Bayes factor to simultaneously provide privacy and utility

Thank you!

BFF4, Harvard University, May 2, 2017