European Research Council
Established by the European Commission

הקרן הלאומית למדע
المؤسّسة الإسرائيلية للعلوم
Israel Science Foundation

Vecteezy

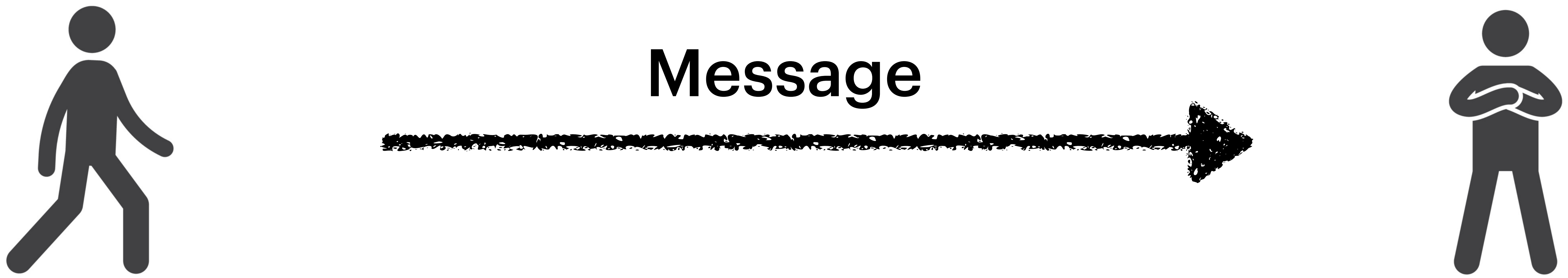
Information Complexity

Dagstuhl seminar 22301

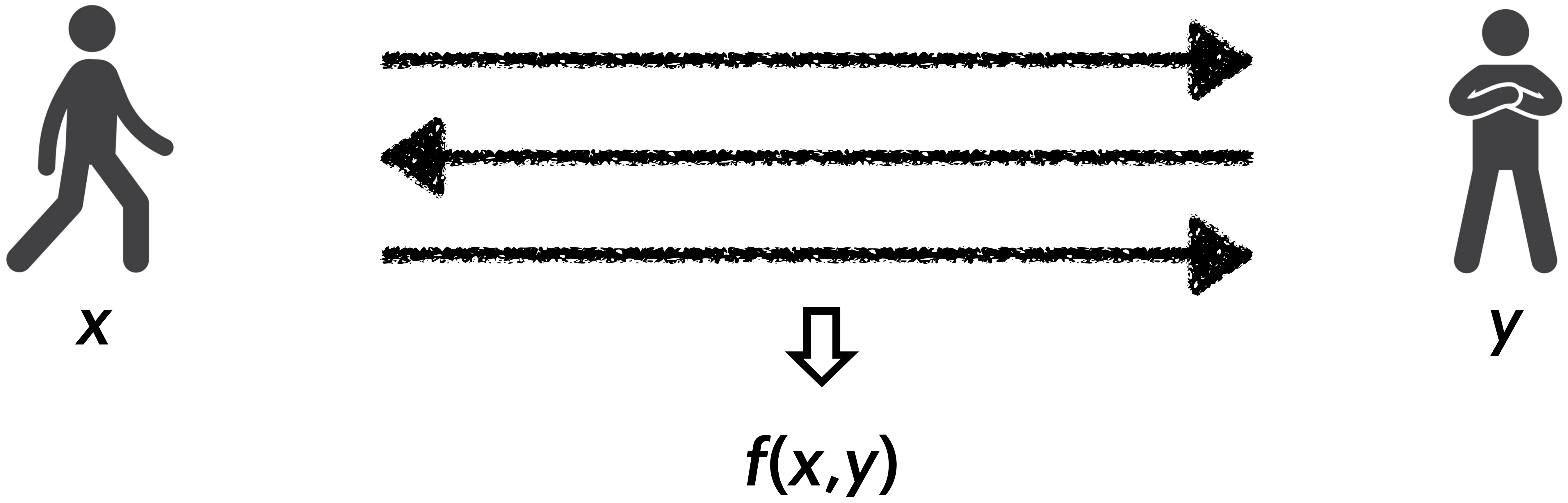
TECHNION | The Henry and Marilyn Taub
Faculty of Computer Science

Yuval Filmus, 28 July 2022

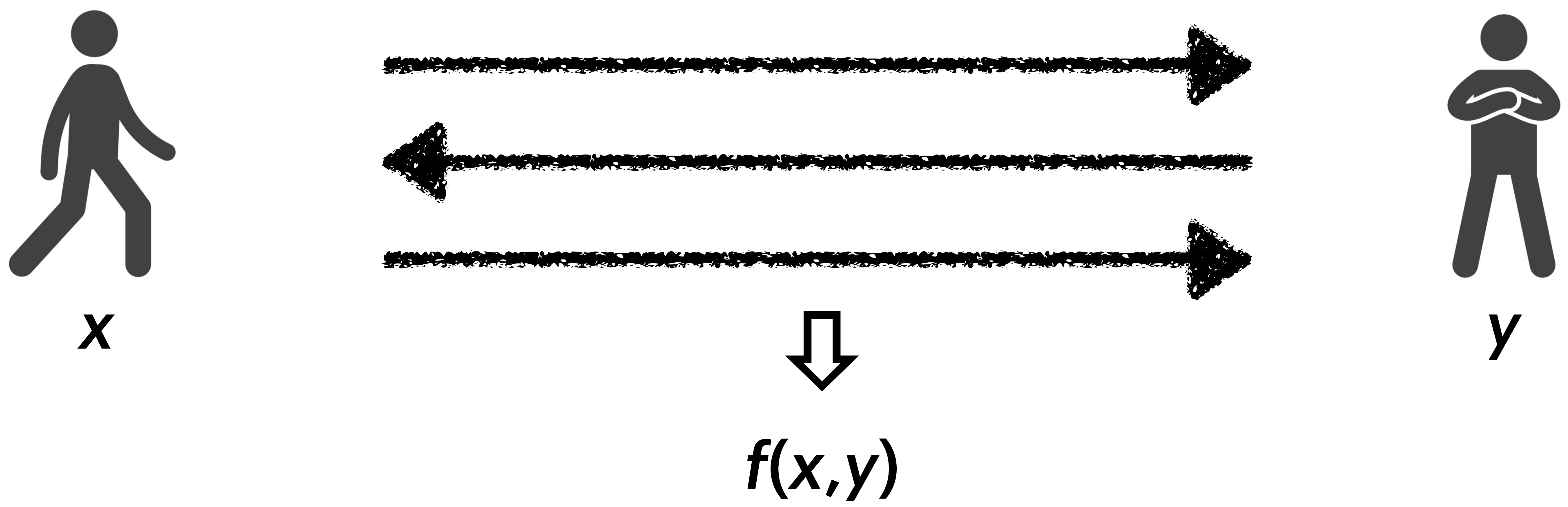
Information Theory



Communication Complexity



Communication Complexity



Variants

Deterministic: output always correct

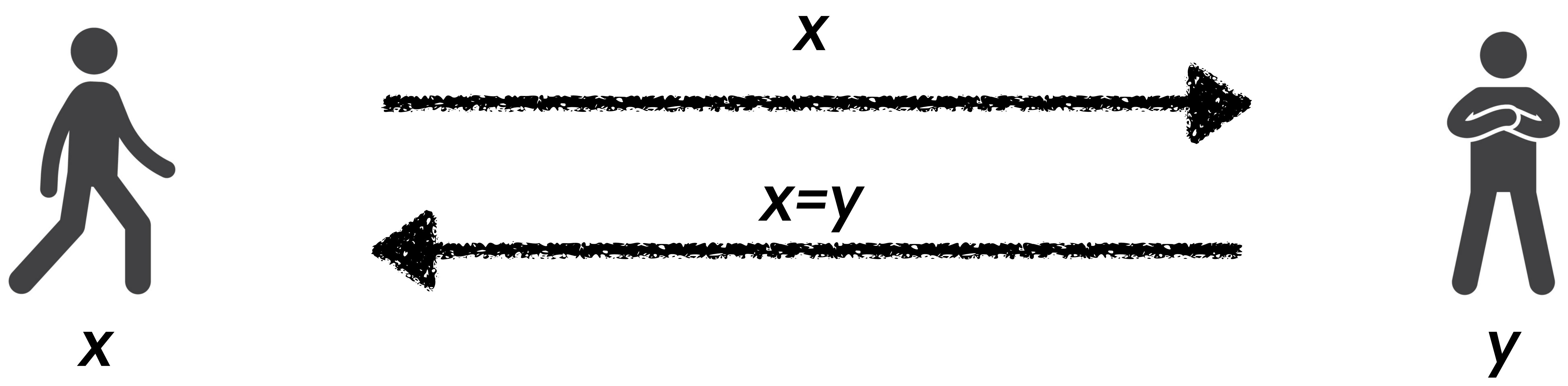
Randomized: output correct w.p. $1-\varepsilon$

Distributional: output correct on $1-\varepsilon$ of inputs

Minimax: randomized = worst distributional

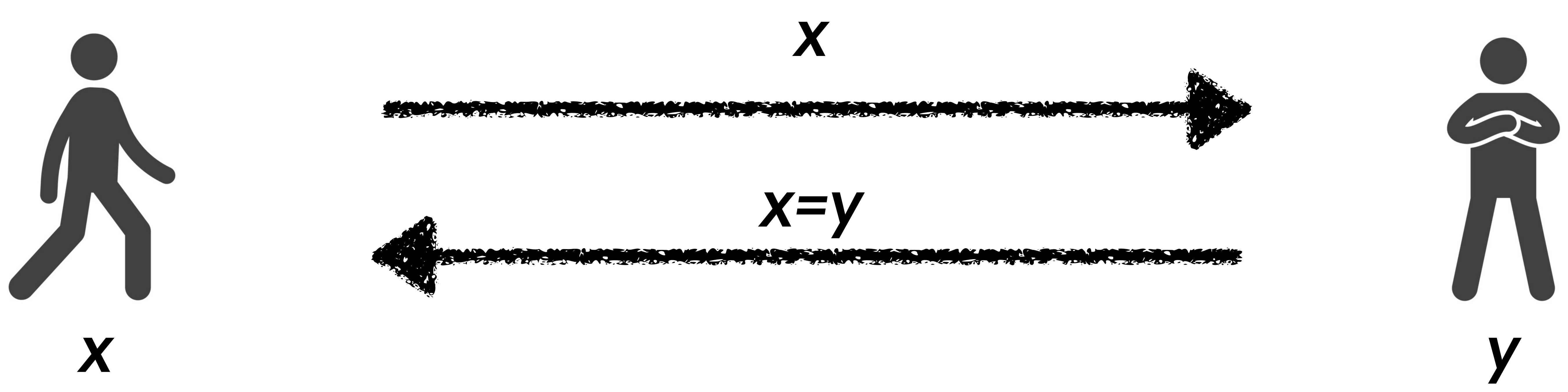
Cost: CC = **maximum** number of bits transmitted

Equality of n-bit strings

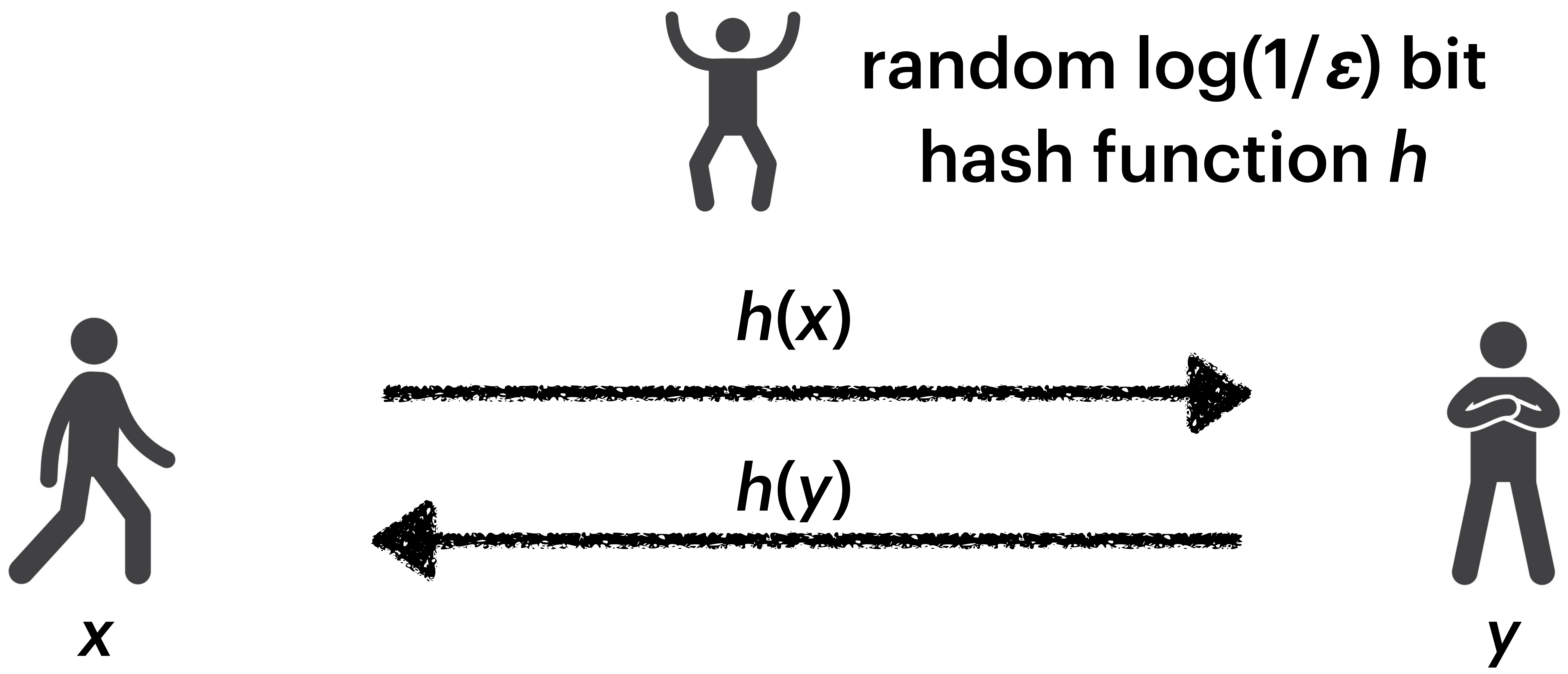


Deterministic: n+1

Equality of n-bit strings



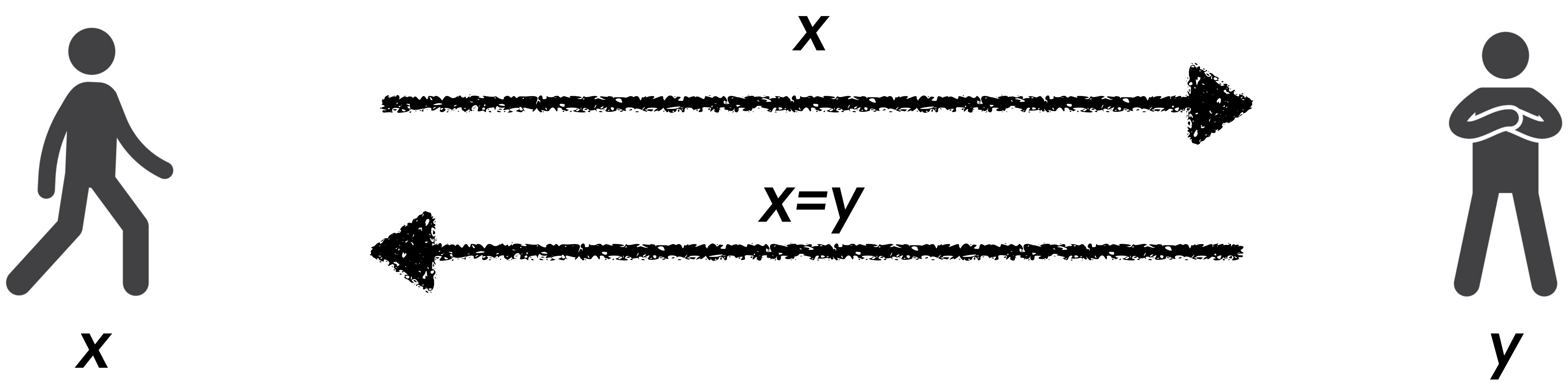
Deterministic: $n+1$



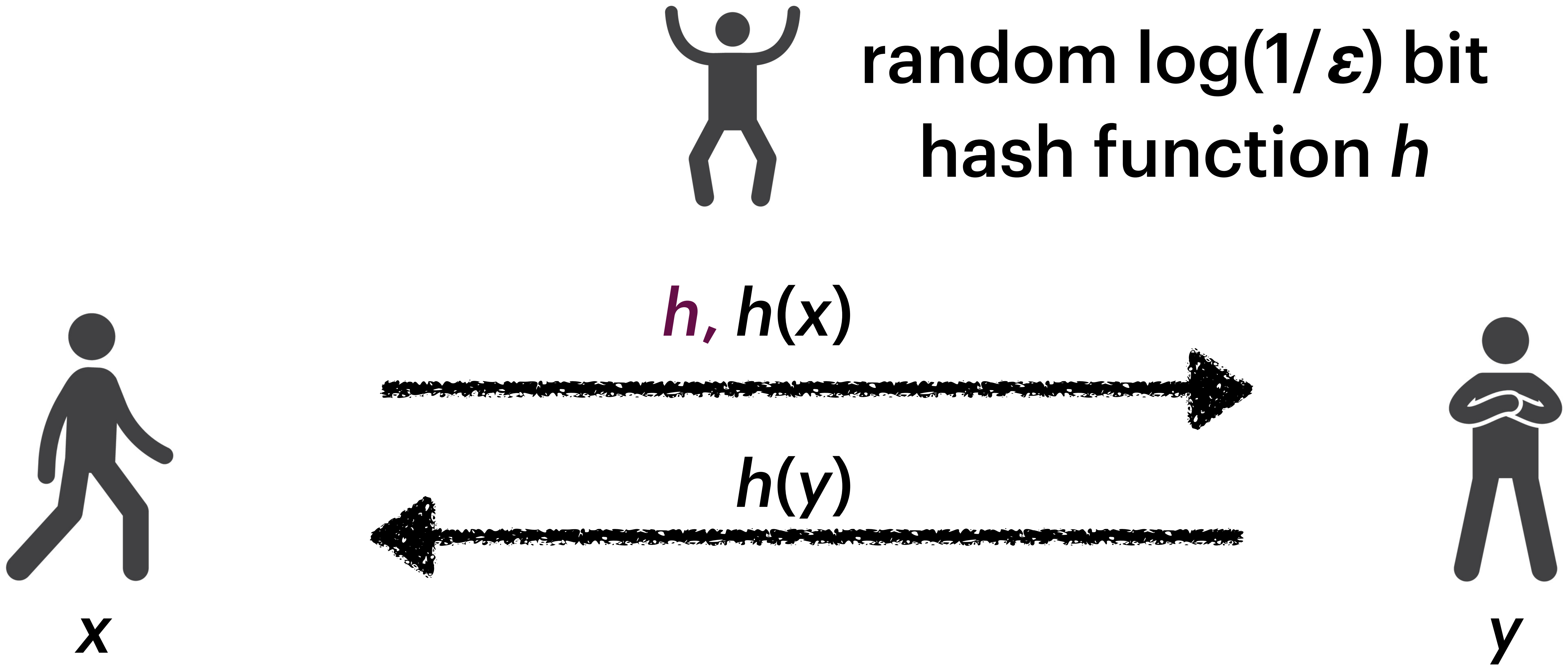
Randomized: $O(\log(1/\varepsilon))$

(with public coins)

Equality of n-bit strings



Deterministic: n+1



Randomized: $O(\log(1/\varepsilon))$

(with public coins)

Randomized: $O(\log(n/\varepsilon))$

(with only private coins)

Greater than on n-bit strings

(randomized, constant ε)

Using binary search, find maximal common prefix of x,y

Following bit reveals which input is larger

Cost per round: $O(1)$

Number of rounds: $\log n$

Total cost: $O(\log n)$

x 01101011

y 01100110

Some hard functions
(randomized)

Inner product: $x_1y_1 \oplus \cdots \oplus x_ny_n$

Randomized cost: $n+1$

Set (non-)disjointness: $x_1y_1 \vee \cdots \vee x_ny_n$

Randomized cost: $\Theta(n)$

Trivial protocol can be improved by constant factor

Why are they hard?
(randomized)

Inner product: $x_1y_1 \oplus \cdots \oplus x_ny_n$

Hard since involves computing n many ANDs

Set (non-)disjointness: $x_1y_1 \vee \cdots \vee x_ny_n$

Hard since involves computing n many ANDs ...

... where answer is almost always 0

How to turn this intuition into a proof?

Direct product
(randomized)

Easier question:

Cost of computing $f(x_1,y_1),\ldots,f(x_n,y_n) \approx n \times$ cost of computing f?

Information theory:

Cost of sending n samples of $X \approx n\,H(X)$

Information complexity:

Cost of computing n copies of $f \approx n\,\mathrm{IC}(f)$

Information complexity

Goal: Cost of computing n copies of $f \approx n\,\mathrm{IC}(f)$

Information complexity of protocol P wrt distribution μ :

$\mathrm{IC}(P,\mu) = \mathrm{I}(\Pi;Y|X) + \mathrm{I}(\Pi;X|Y)$, where X,Y=inputs, Π=transcript of P

“What Alice learns about Bob’s input from transcript” +
“What Bob learns about Alice’s input from transcript”

Information complexity

Goal: Cost of computing n copies of $f \approx n\,\mathrm{IC}(f)$

Information complexity of protocol P wrt distribution μ:

$\mathrm{IC}(P,\mu) = \mathrm{I}(\Pi;Y|X) + \mathrm{I}(\Pi;X|Y)$, where X, Y=inputs, Π=transcript of P

"What Alice learns about Bob's input from transcript" +
"What Bob learns about Alice's input from transcript"

IC of function f wrt distribution μ and error ε:

$\mathrm{IC}(f,\mu,\varepsilon) = \min \mathrm{IC}(P,\mu)$ over all P computing f with error ε wrt μ

Information complexity

Goal: Cost of computing n copies of $f \approx n\,\mathrm{IC}(f)$

Information complexity of protocol P wrt distribution μ:

$\mathrm{IC}(P,\mu) = I(\Pi;Y|X) + I(\Pi;X|Y)$, where X,Y=inputs, Π=transcript of P

"What Alice learns about Bob's input from transcript" +
"What Bob learns about Alice's input from transcript"

IC of function f wrt distribution μ and error ε:

$\mathrm{IC}(f,\mu,\varepsilon) = \min \mathrm{IC}(P,\mu)$ over all P computing f with error ε wrt μ

IC of function f with error ε:

$\mathrm{IC}(f,\varepsilon) = \max \mathrm{IC}(f,\mu,\varepsilon)$ over all distributions μ

Properties of information complexity

$IC(P,\mu) = I(\Pi;Y|X) + I(\Pi;X|Y)$, where X, Y=inputs, Π=transcript of P

$IC(f,\mu,\varepsilon)$ = min $IC(P,\mu)$ over all P computing f with error ε wrt μ

$IC(f,\varepsilon)$ = max $IC(f,\mu,\varepsilon)$ over all distributions μ

IC lower bounds communication: $IC(f,\boldsymbol{\mu},\varepsilon)\leq CC(f,\boldsymbol{\mu},\varepsilon)$

Direct product: $IC(f\otimes g,\mu\otimes\nu,\varepsilon^*)=IC(f,\mu,\varepsilon)+IC(g,\nu,\varepsilon)$

"Source coding theorem": $CC(f^n,\mu^n,\varepsilon^*)\approx n\ IC(f,\mu,\varepsilon)$

*error per copy

Properties of information complexity

$IC(P,\mu) = I(\Pi;Y|X) + I(\Pi;X|Y)$, where X, Y=inputs, Π=transcript of P

$IC(f,\mu,\varepsilon) = \min IC(P,\mu)$ over all P computing f with error ε wrt μ

$IC(f,\varepsilon) = \max IC(f,\mu,\varepsilon)$ over all distributions μ

IC lower bounds communication: $IC(f,\mu,\varepsilon) \leq CC(f,\mu,\varepsilon)$

Direct product: $IC(f \otimes g, \mu \otimes \nu, \varepsilon^*) = IC(f,\mu,\varepsilon) + IC(g,\nu,\varepsilon)$

"Source coding theorem": $CC(f^n,\mu^n,\varepsilon^*) \approx n\, IC(f,\mu,\varepsilon)$

No analog of Shannon–Fano:

Gap between IC and CC can be exponential!

(True even when measuring *average* number of bits communicated)

*error per copy

Exact complexity of set disjointness

"Source coding theorem": $CC(f^n,\mu^n,\varepsilon^*)\approx n\ IC(f,\mu,\varepsilon)$

Version for OR: $CC(\vee \text{ of } n \text{ copies of } f, o(1))\approx n\ IC^0(f,0)$

where $IC^0(f,0)=\max IC(f,\mu,0)$ over μ supported on $f^{-1}(0)$

Example: $IC^0(AND,0) = 0.4827\ldots$

Conclusion: $CC(\text{set-disjointness},o(1))\approx 0.4827\ldots\ n$

No explicit protocol is known!

Buzzer protocol

Optimal protocol for AND (for symmetric distributions)

Alice gets a bit x, Bob gets a bit y

Alice chooses a random $t_a \in [0,1]$

Bob chooses a random $t_b \in [0,1]$

Buzzer protocol

Optimal protocol for AND (for symmetric distributions)

Alice gets a bit x, Bob gets a bit y

Alice chooses a random $t_a \in [0,1]$

Bob chooses a random $t_b \in [0,1]$

A timer counts from 0 to 1 continuously

Buzzer protocol

Optimal protocol for AND (for symmetric distributions)

Alice gets a bit x, Bob gets a bit y

Alice chooses a random $t_a \in [0,1]$

Bob chooses a random $t_b \in [0,1]$

A timer counts from 0 to 1 continuously

$\begin{cases} \text{At time } t_a\text{: if } x=0\text{, Alice presses buzzer, protocol outputs 0} \\ \text{At time } t_b\text{: if } y=0\text{, Bob presses buzzer, protocol outputs 0} \end{cases}$

Buzzer protocol

Optimal protocol for AND (for symmetric distributions)

Alice gets a bit x, Bob gets a bit y

Alice chooses a random $t_a \in [0,1]$

Bob chooses a random $t_b \in [0,1]$

A timer counts from 0 to 1 continuously

$$\begin{cases} \text{At time } t_a\text{: if } x=0\text{, Alice presses buzzer, protocol outputs 0} \\ \text{At time } t_b\text{: if } y=0\text{, Bob presses buzzer, protocol outputs 0} \end{cases}$$

At time 1: protocol outputs 1

Generalized protocols

The buzzer protocol is not a real protocol!

It can be discretized to a real protocol with r rounds whose information complexity is OPT+$\Theta(1/r^2)$.

OPT cannot be achieved using any real protocol!

Challenge:

Define a generalized notion of protocols which achieves the optimal information complexity **exactly** for every f.

More open questions

Amortized communication complexity for zero error?

Information complexity for multiple parties?

Is $\mathrm{CC}(f)$ polynomial in $\mathrm{IC}(P) \log \mathrm{CC}(P)$?

Bibliography

Monographs

Anup Rao and Amir Yehudayoff, *Communication Complexity: And Applications*, 2020

Surveys

Mark Braverman, *Communication and information complexity*, Proc. ICM 2022

Omri Weinstein, *Information Complexity and the Quest for Interactive Compression*, 2015

Papers

Barak, Braverman, Chen, Rao, *How to compress interactive communication*, 2013

Braverman, Rao, *Information equals amortized communication*, 2014

Braverman, Garg, Pankratov, Weinstein, *From information to exact communication*, 2013

2022 Abacus Medal awarded to Mark Braverman