On the Measurement of Information

Abby Wanyu Liu

LTCI, Telecom ParisTech, Université Paris-Saclay, Paris, France

LRI, Univ. Paris-Sud, CNRS, Inria, Université Paris-Saclay, Orsay, France

About Me

- Last year Ph.D. student in Paris
- Working with 
- Participant of the 2nd summer school
- Keywords: HCI, information theory, mutual information, bayesian, computational interaction

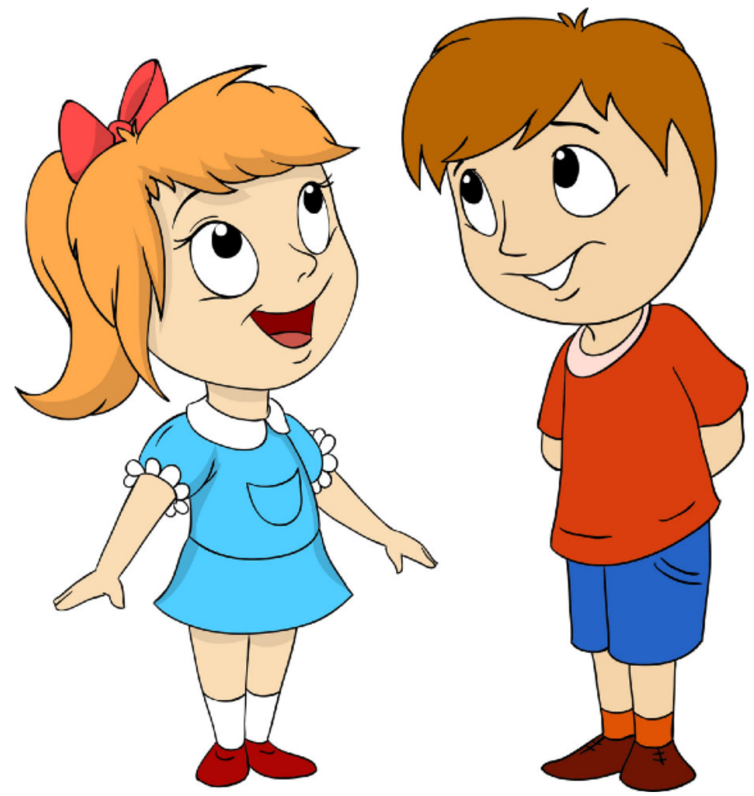
Agenda

* Information in human-computer interaction
* Information Theory to quantify information
* Bayesian Information Gain (BIG)
* Information-theoretic measures to characterize interaction

Slides and Notebook exercises can be found on:

https://github.com/wanyuliu/4thComputationalInteraction

- Information is everywhere

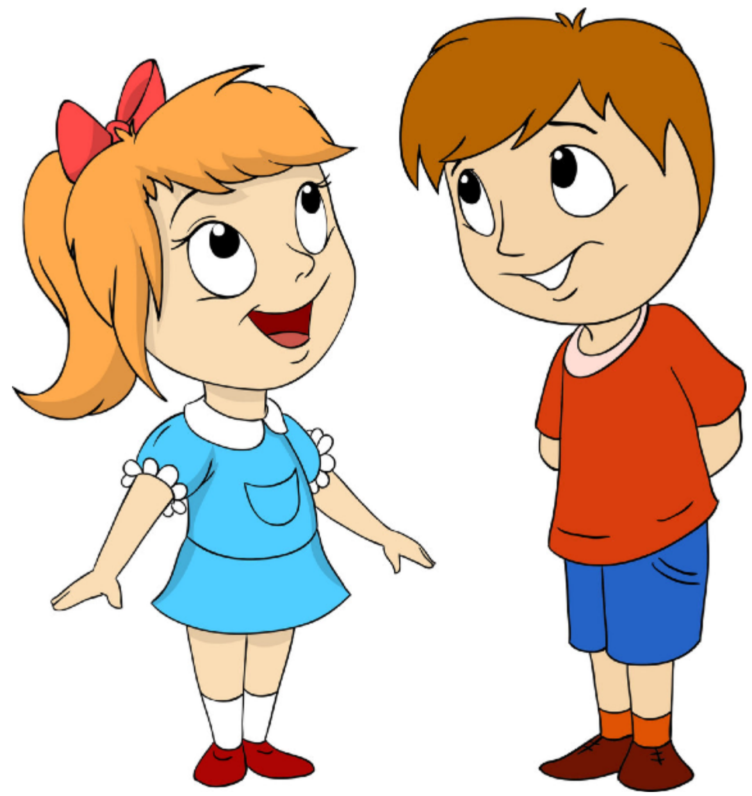


Jean-Pierre: Where do we go for dinner tonight? French or Fish&Chips?

Chantal: Let’s have Fish&Chips.

Information

- Information is everywhere



Jean-Pierre: Where do we go for dinner tonight? French or Fish&Chips?

Chantal: I don't know. We had Fish&Chips the other day so I would feel more like French tonight. But I really like Fish&Chips. Btw Louise asked us if we want join her for dinner at Vincent's place. Vincent just came back from Sri Lanka and took some cooking courses there. But I'm not big fan of spicy food so I don't know. What do you reckon?

No Information !

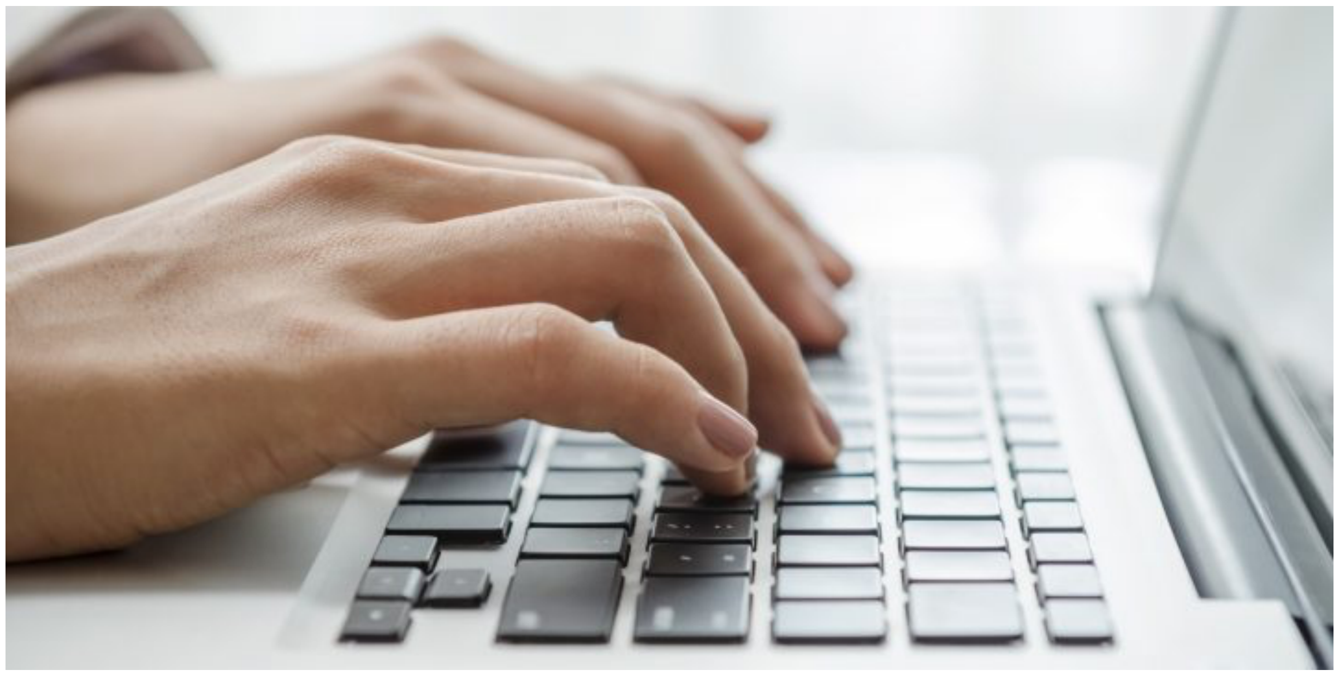
- Information is everywhere

Information is any entity or form that provides the answer to a **question** of some kind or resolves **uncertainty**.

Source: https://en.wikipedia.org/wiki/Information

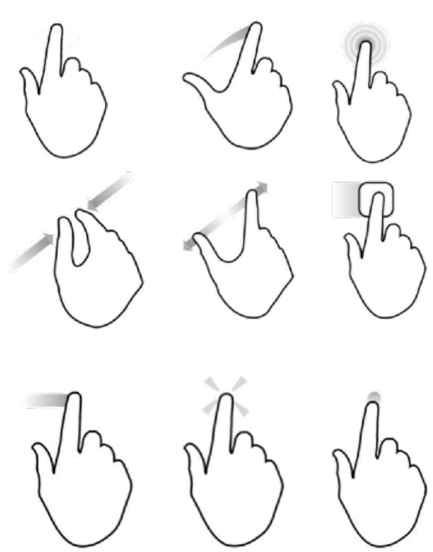
- Information in human-computer interaction

Characters

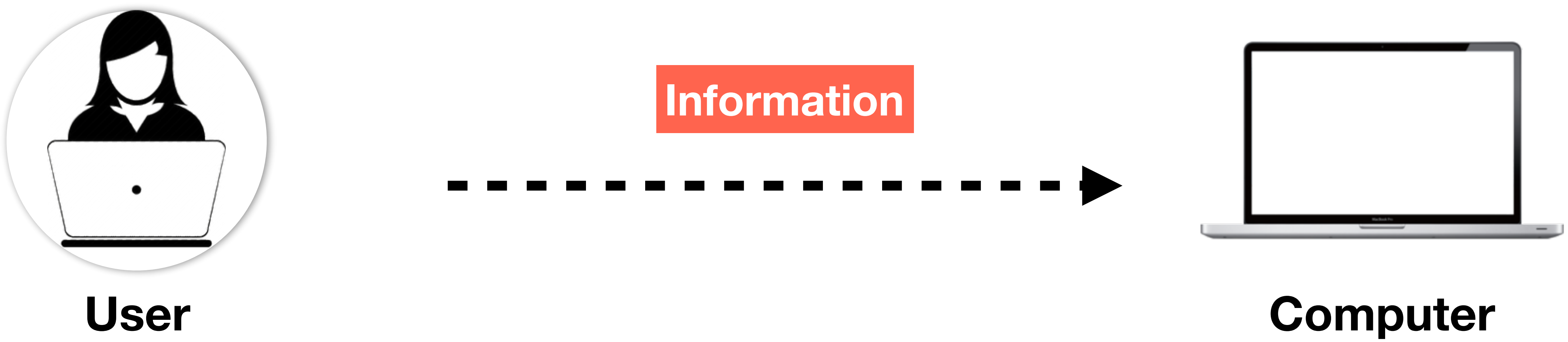


Words

Gestures



- Information in human-computer interaction

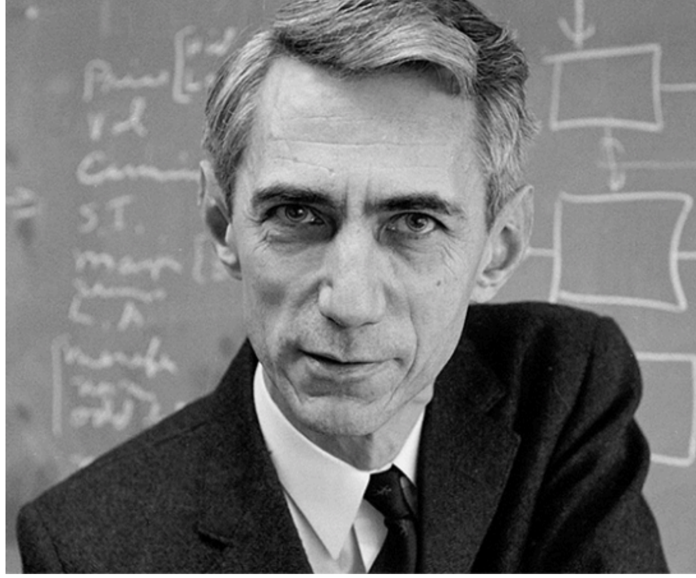


Can we quantify information?

Information in a more general sense?

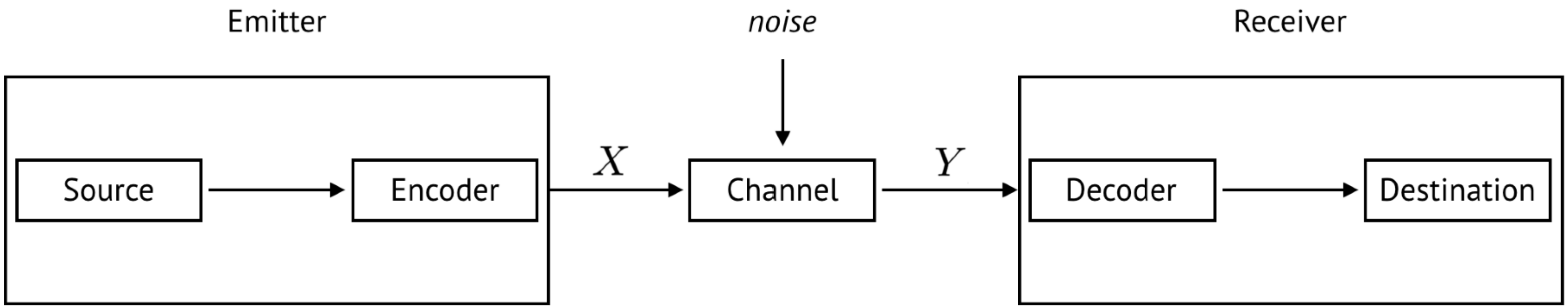
HCI as information transmission in *What Is Interaction?* (Kasper Hornbæk & Antti Oulasvirta, CHI'17)

- Information Theory to quantify information



Claude Shannon

A mathematical theory of communication (1948)



Elements of information theory. Cover, T. M., & Thomas, J. A. (2012).

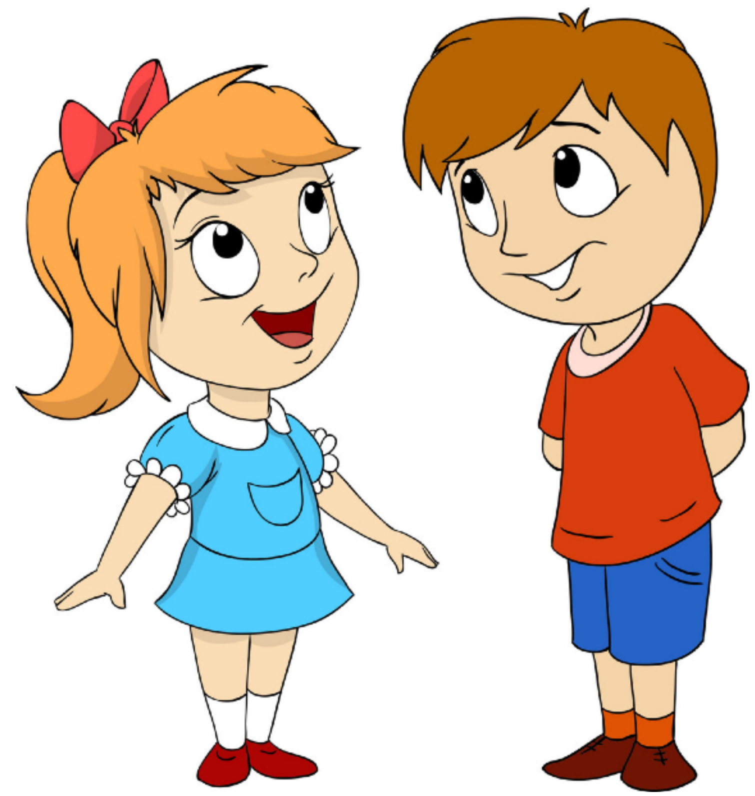
- Information Theory to quantify information

Random variable *X* **can take values in** $\{x_1, x_2, \ldots, x_n\}$

$P_1 \quad P_2 \quad \ldots \quad P_n$

Entropy: $H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i$ $\qquad 0 \le H(X) \le \log N$

- Information Theory to quantify information



Jean-Pierre: Where do we go for dinner tonight? French or Fish&Chips?

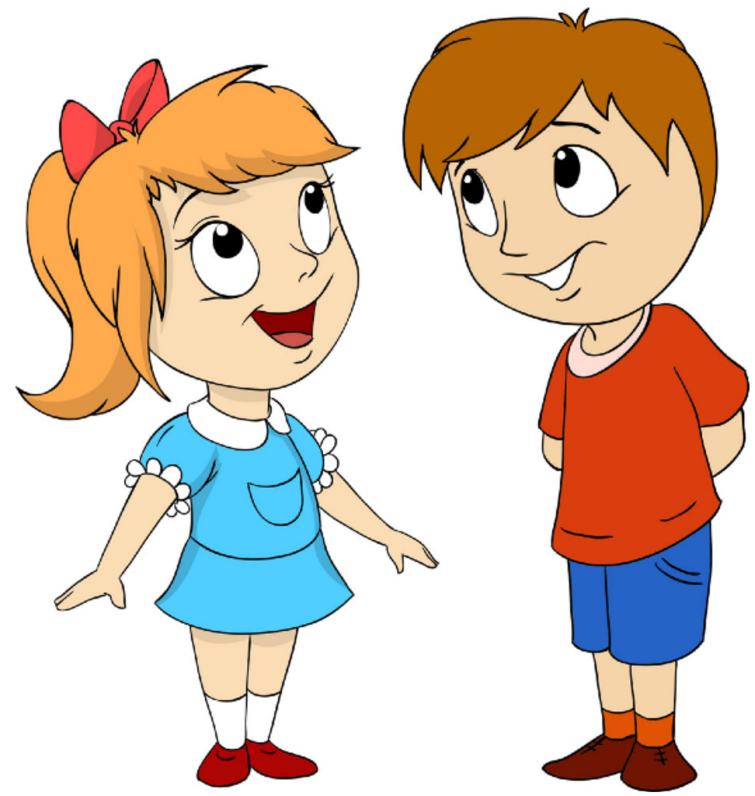
Chantal: Let's have Fish&Chips.

$\mathbf{x_1 = 0}$ $\mathbf{x_2 = 1}$

Before: H(X) = log 2 = 1 bit

After: H(X)' = 0

- Information Theory to quantify information



Jean-Pierre: Where do we go for dinner tonight? French or Fish&Chips?

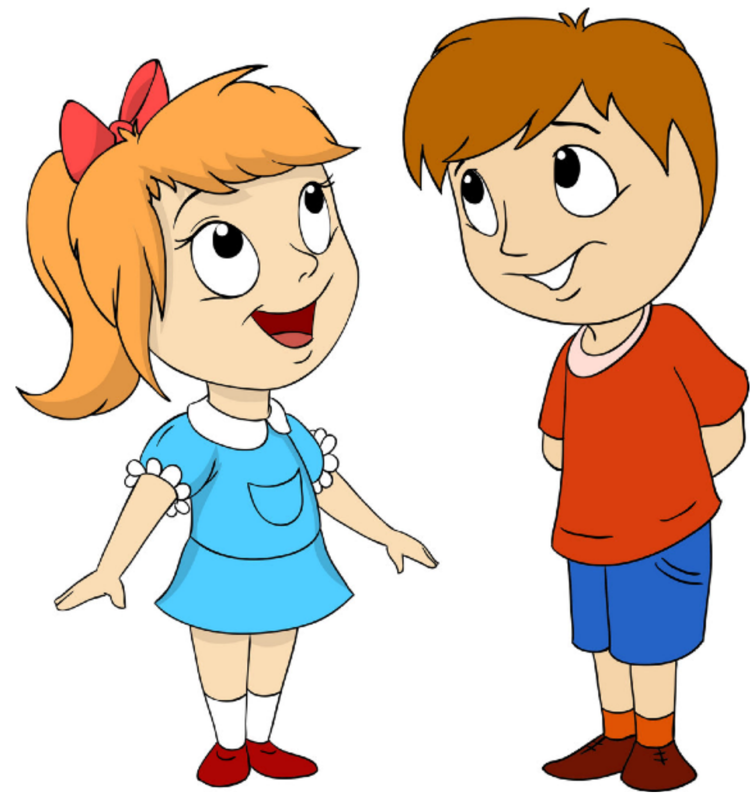
Chantal: Let's have French.

$x_1 = 1$ $x_2 = 0$

Before: H(X) = 0.72 bit

After: H(X)' = 0

- Information Theory to quantify information

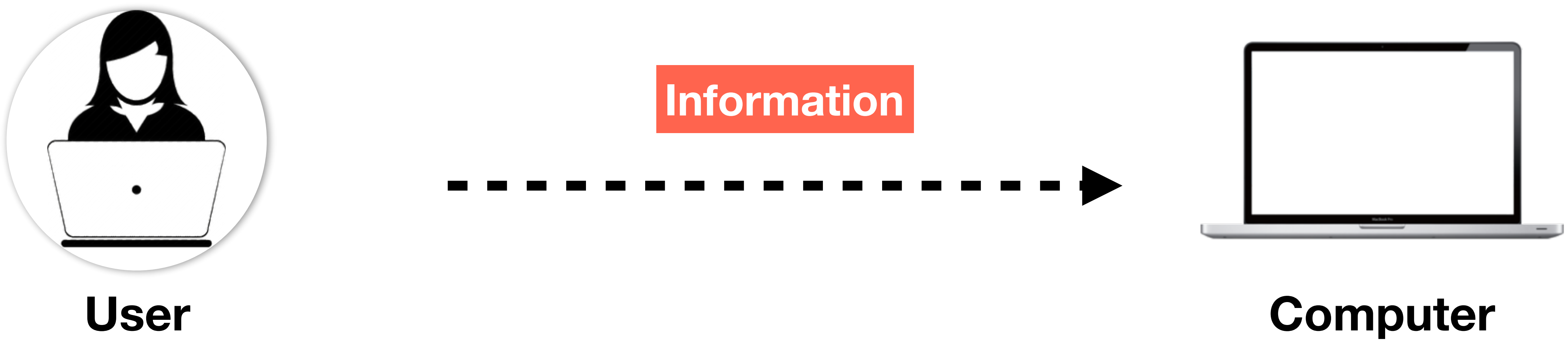


Jean-Pierre: Where do we go for dinner tonight? French or Fish&Chips?

Chantal: I don't know. We had Fish&Chips the other day so I would feel more like French tonight. But I really like Fish&Chips. Btw Louise asked us if we want join her for dinner at Vincent's place. Vincent just came back from Sri Lanka and took some cooking courses there. But I'm not big fan of spicy food so I don't know. What do you reckon?

There is still uncertainty !

- Information Theory to quantify information



Characters

Words

Gestures

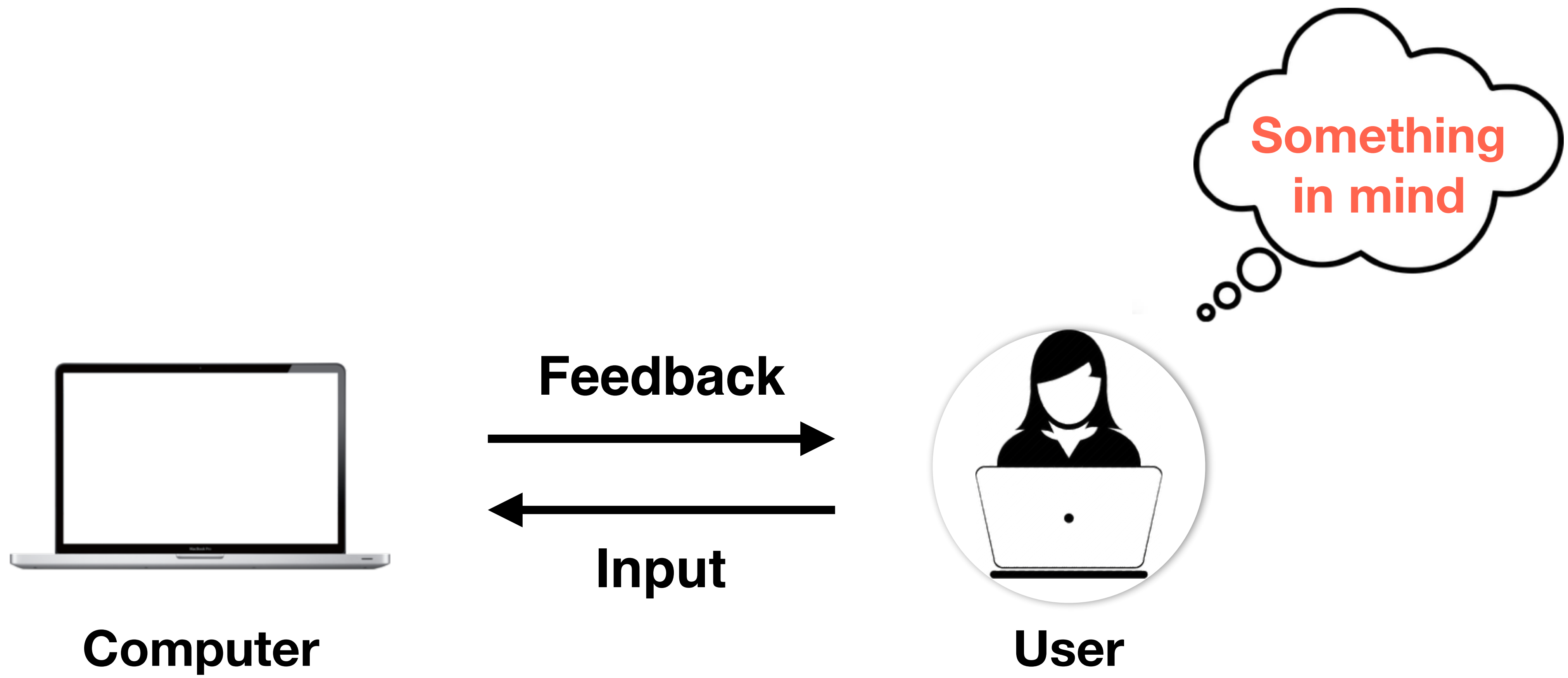
Commands

Any input?

- Bayesian Information Gain (BIG)

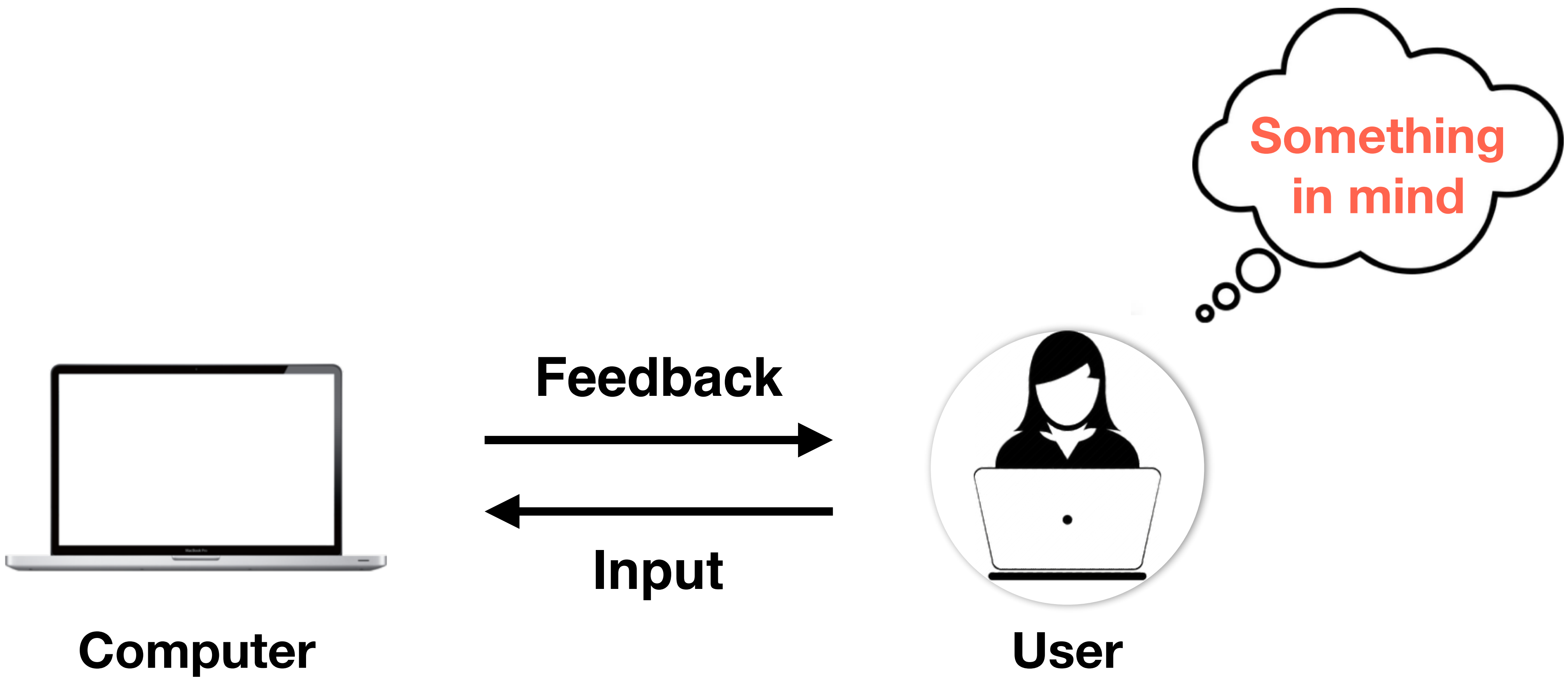
The computer gains **information** from the **user input** to reduce its **uncertainty** about the user's goal.

- Bayesian Information Gain (BIG)



- To look for a restaurant
- To type a word
- To draw a gesture
- To select an icon
- To do something
- ….

- Bayesian Information Gain (BIG)



- Uncertainty about this something
- Uncertainty reduces gradually when receiving user input

- Bayesian Information Gain (BIG)

  - **Executes the user input only** Multiscale navigation
  - **Maximizes the expected information gain** $IG(\Theta|X=x,Y)$ BIGnav
  - **Leverages the expected information gain** BIGFile

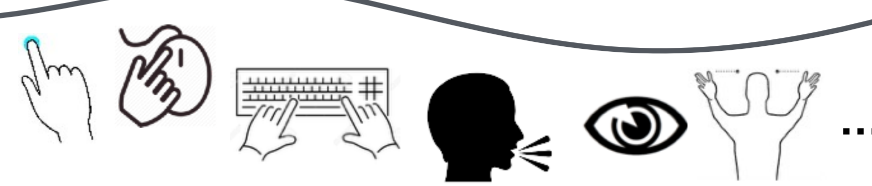
Feedback $X = x$

Computer

- **Prior knowledge:** $P(\Theta = \theta)$
- **User behavior model:** $P(Y=y|\Theta=\theta, X=x)$
- **Update knowledge:** $P(\Theta=\theta \mid X=x, Y=y)$
- **Calculate Information Gain:** $IG(\Theta|X=x, Y=y)$

User

Input $Y = y$



- Bayesian Information Gain (BIG)

- **The computer's Uncertainty about the user's goal**

$$H(\Theta) = -\sum_{i=1}^{n} P(\Theta = \theta_i) \log_2 P(\Theta = \theta_i)$$

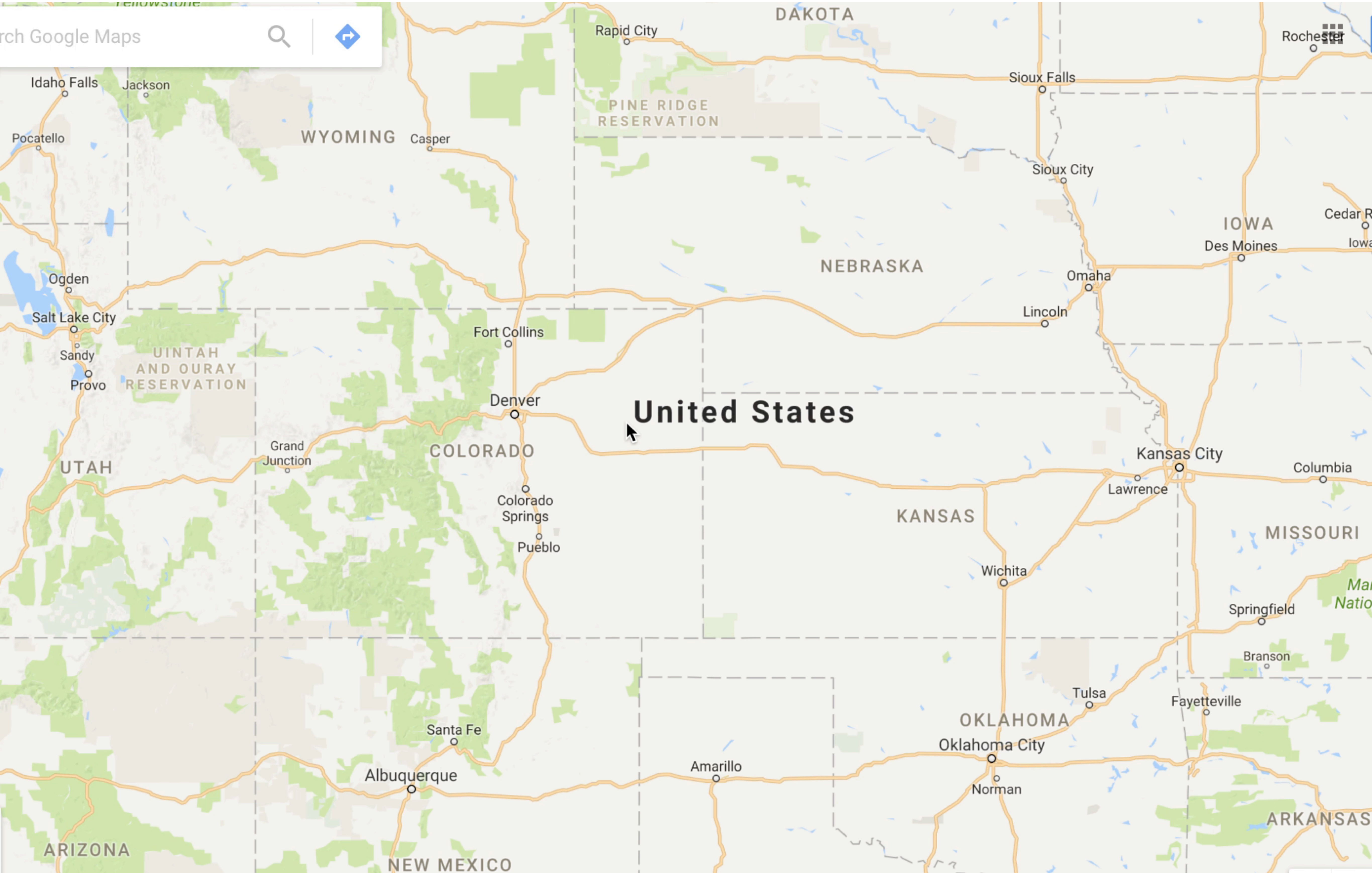
- **The computer's updated knowledge about the user's goal**

$$P(\Theta = \theta \mid X = x, Y = y) = \frac{P(Y = y \mid \Theta = \theta, X = x) P(\Theta = \theta)}{P(Y = y \mid X = x)}$$

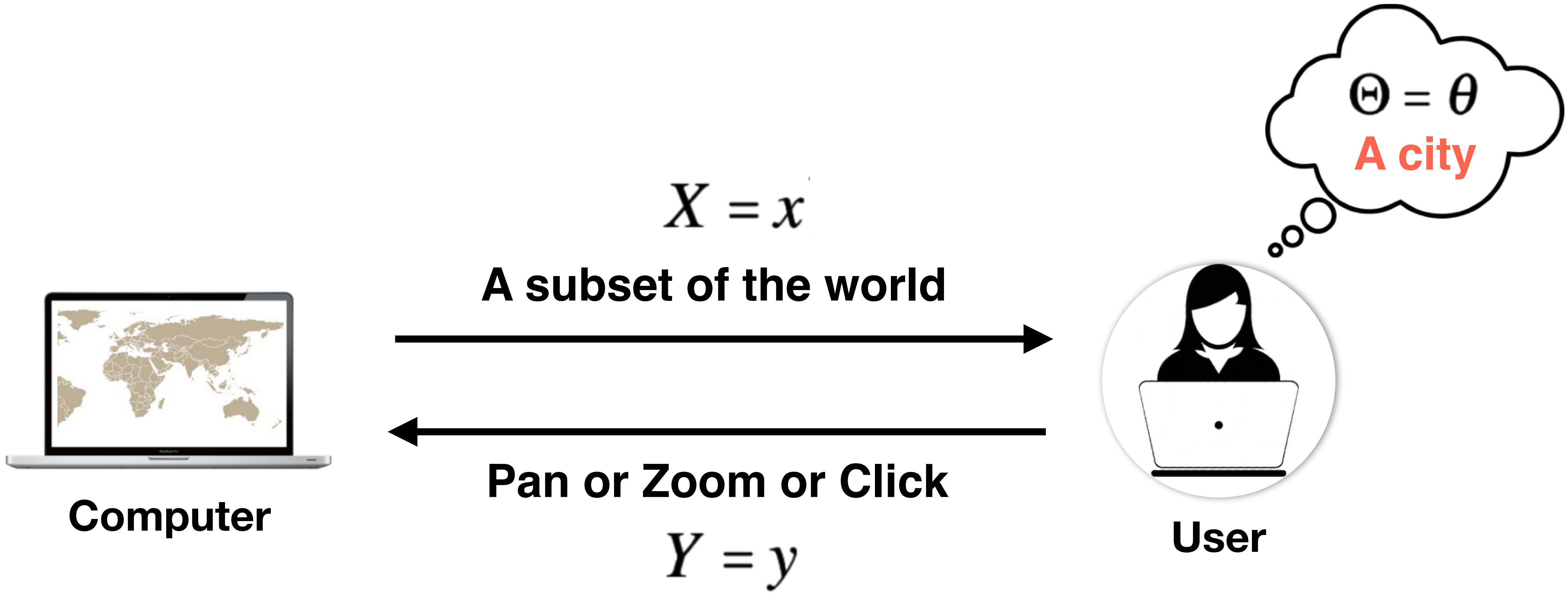
- **The information in the user's input for reducing the computer's uncertainty**

$$IG(\Theta \mid X = x, Y = y) = H(\Theta) - H(\Theta \mid X = x, Y = y)$$

- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation

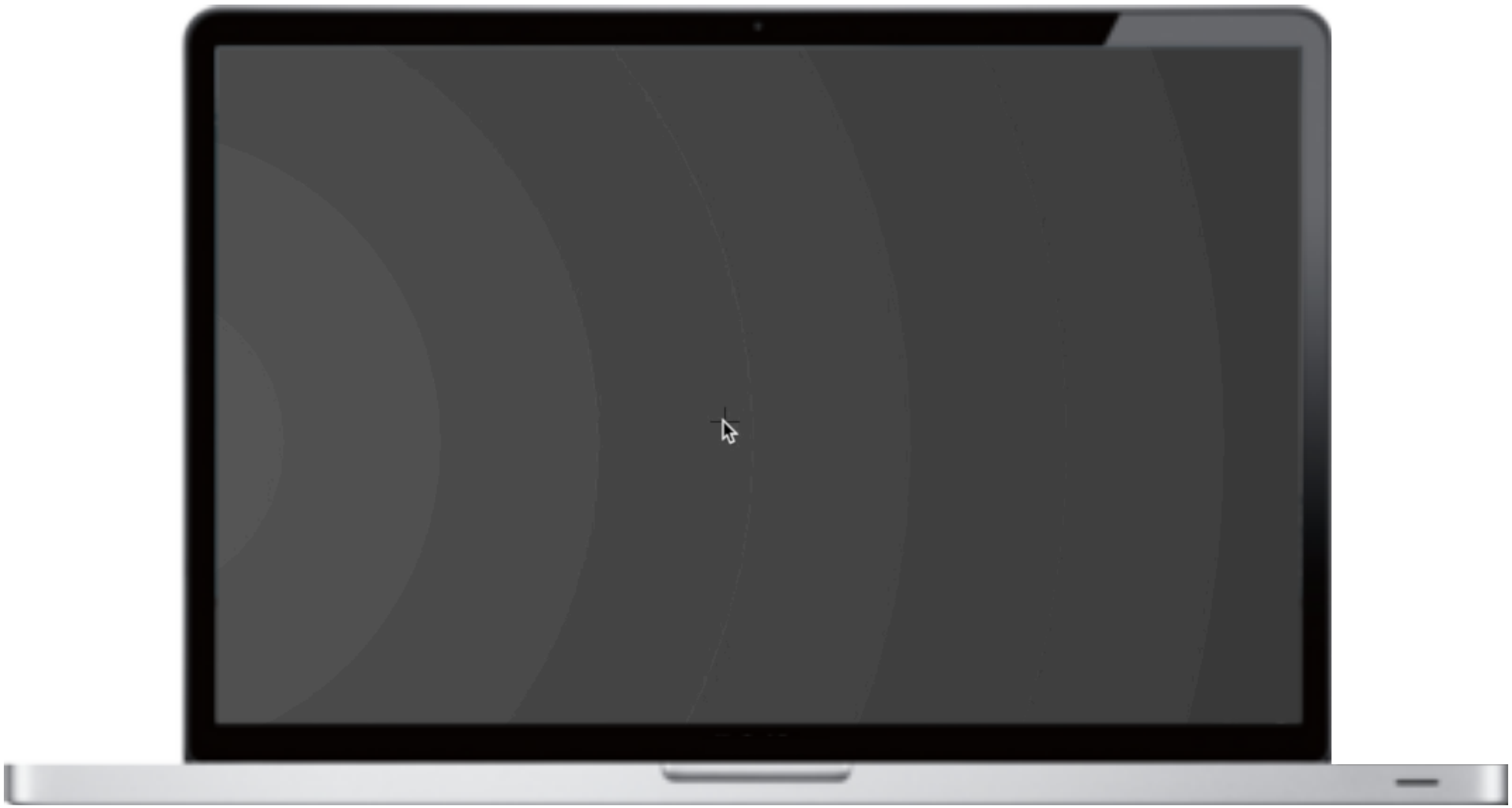


$P(\Theta = \theta)$ **The computer's prior knowledge about the user's goal**

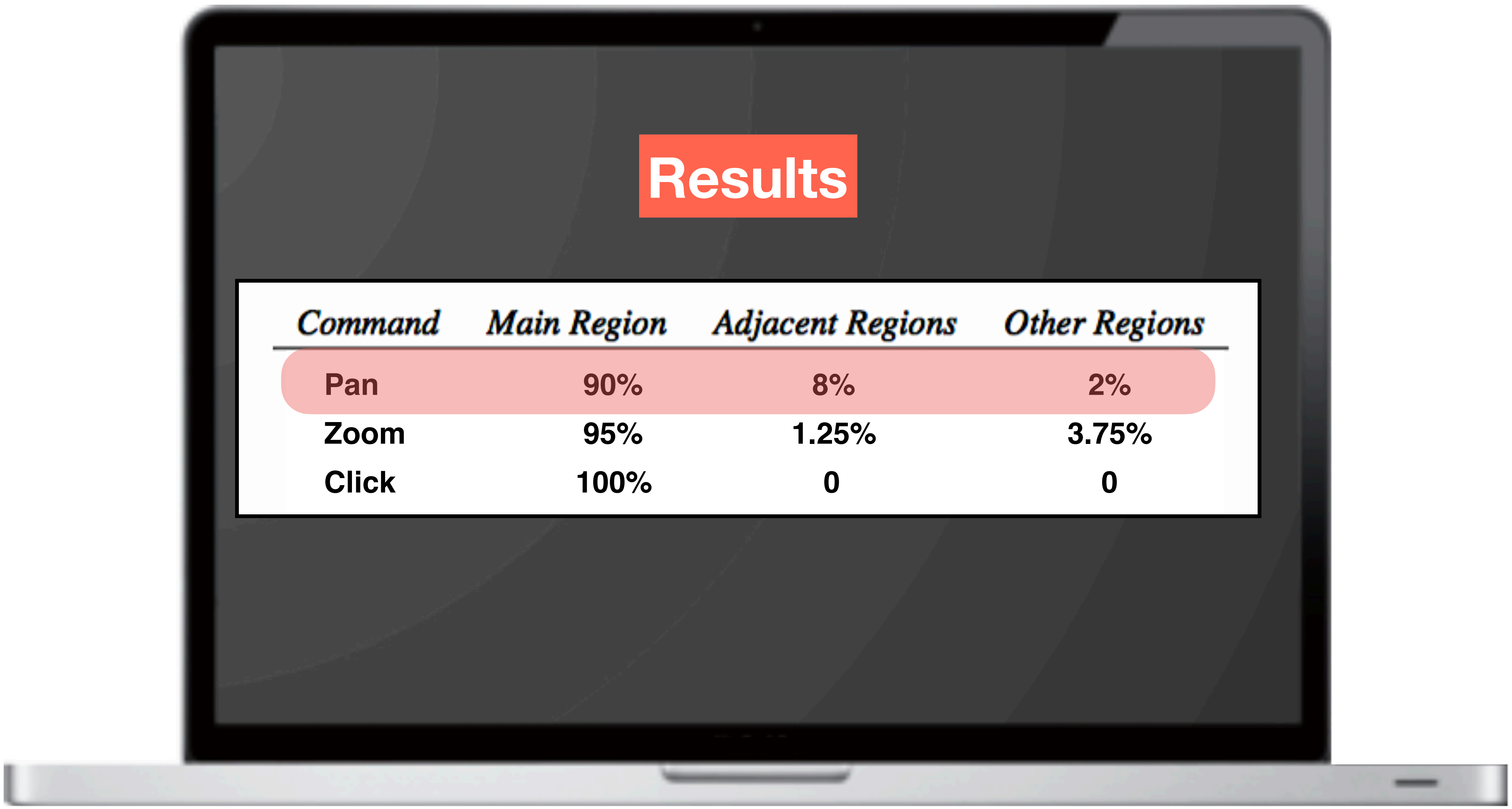
$P(Y = y | \Theta = \theta, X = x)$ **User behavior model**

- Bayesian Information Gain (BIG) - Standard navigation

- **A calibration session to understand user behavior** $P(Y = y|\Theta = \theta, X = x)$



- Bayesian Information Gain (BIG) - Standard navigation

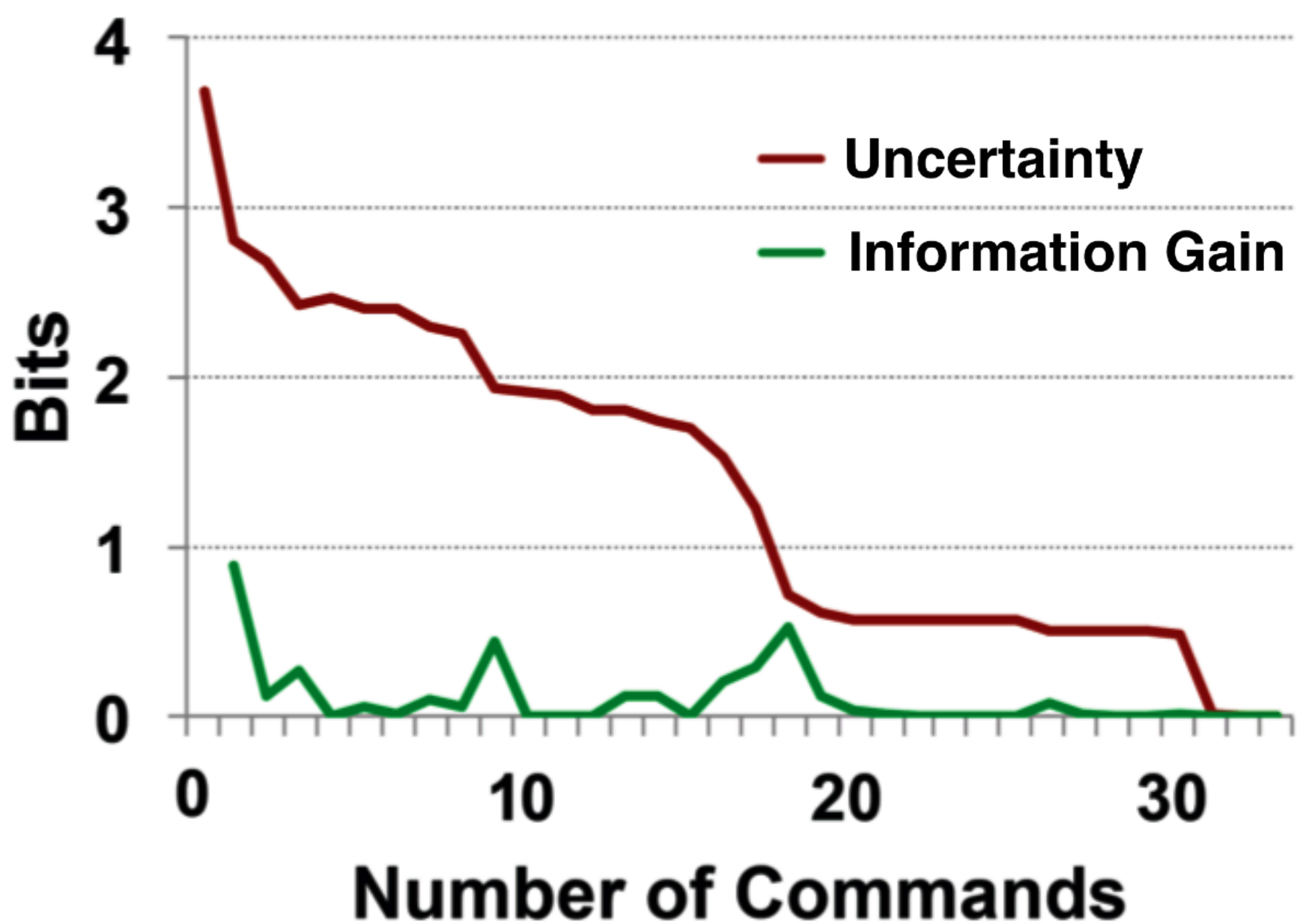


Command	*Main Region*	*Adjacent Regions*	*Other Regions*
Pan	90%	8%	2%
Zoom	95%	1.25%	3.75%
Click	100%	0	0

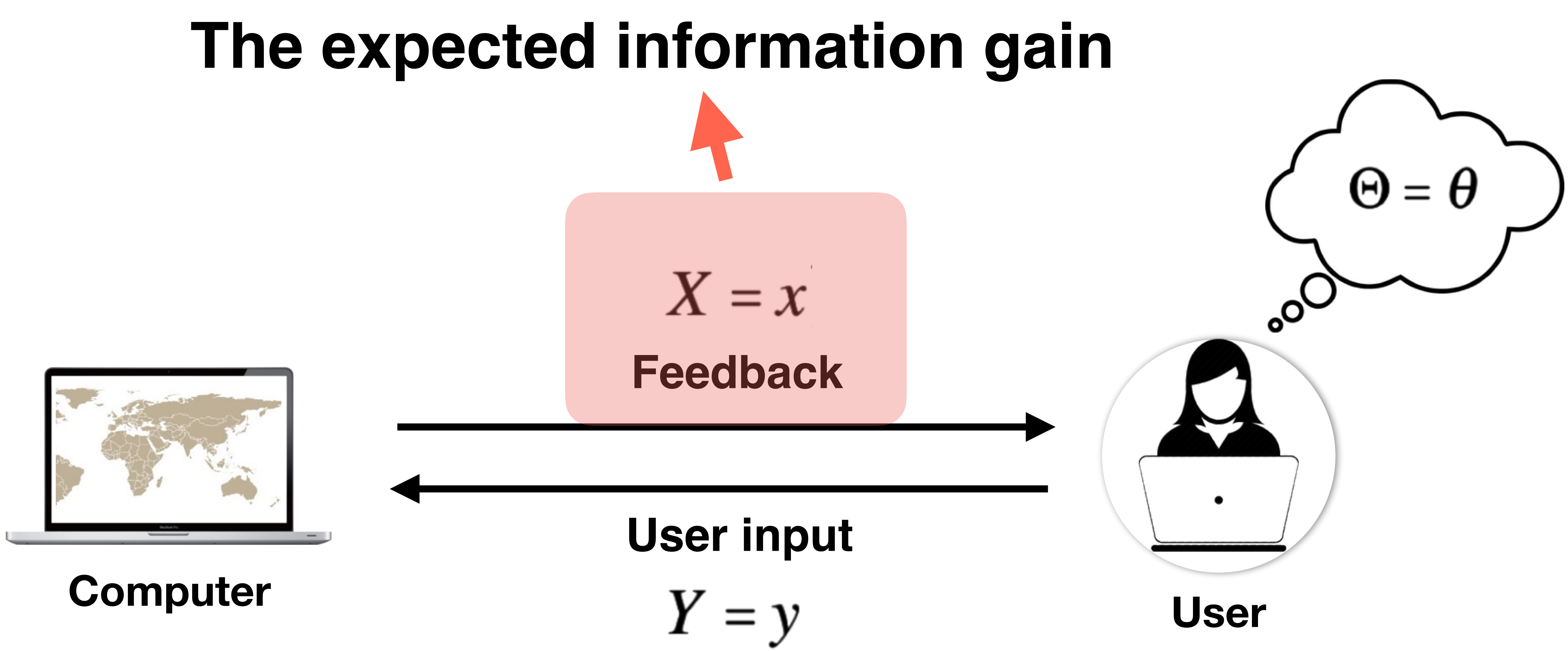
- Bayesian Information Gain (BIG) - Standard navigation

  - **Executes the user input only** **Multiscale navigation**

  - **Each user input does not provide much information for the computer to know her goal**

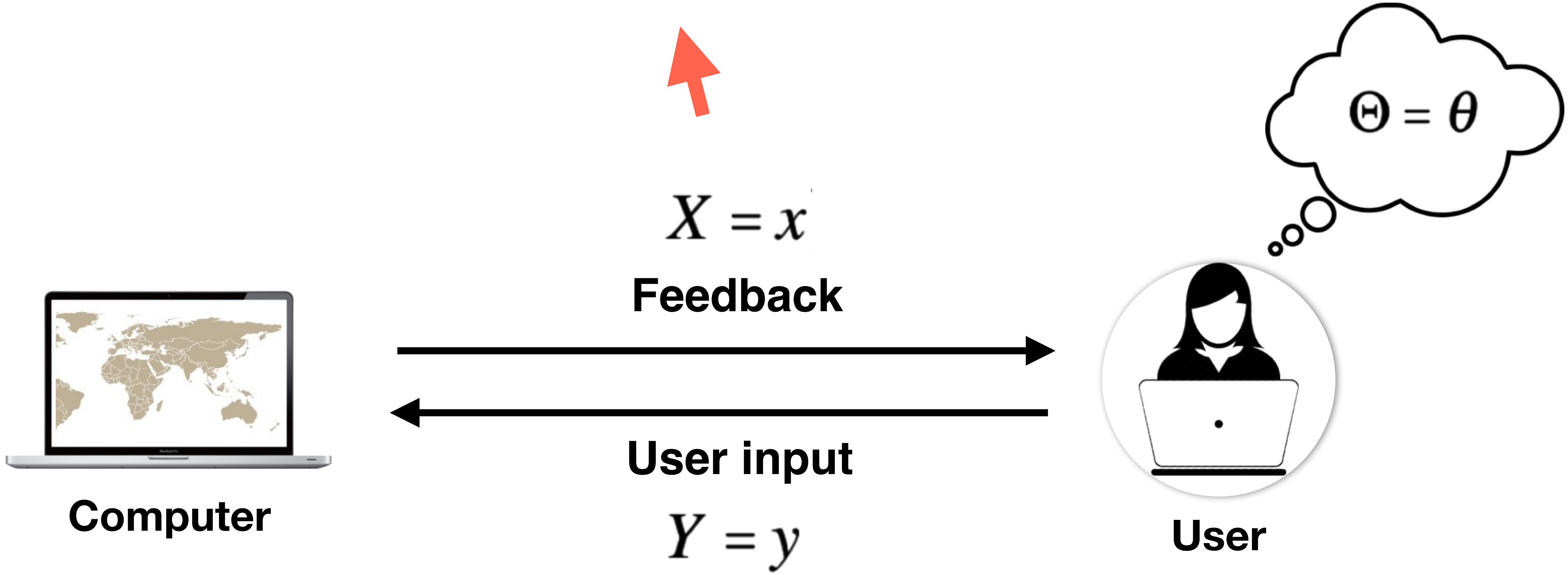


Can we challenge users to give more information?



- Bayesian Information Gain (BIG) - BIGnav
  - **Maximizes the expected information gain** $IG(\Theta|X = x, Y)$

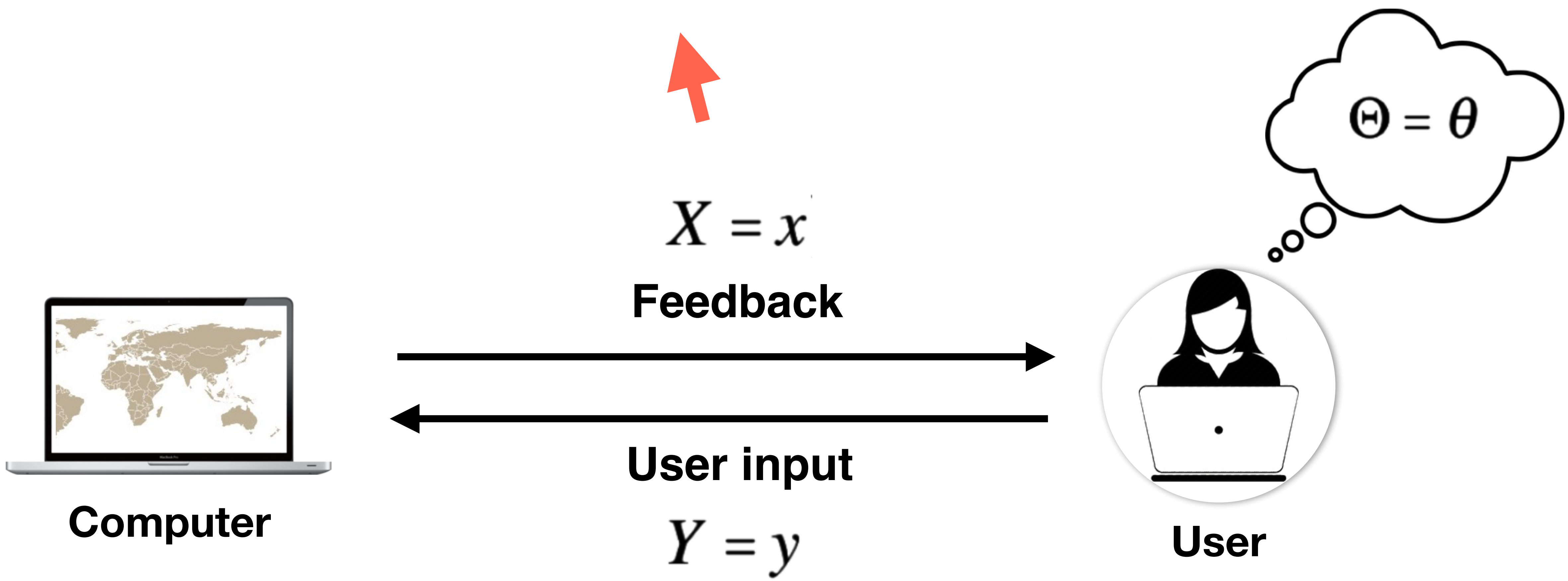
Choose the feedback (a view) that maximizes the expected information gain from the user's subsequent input



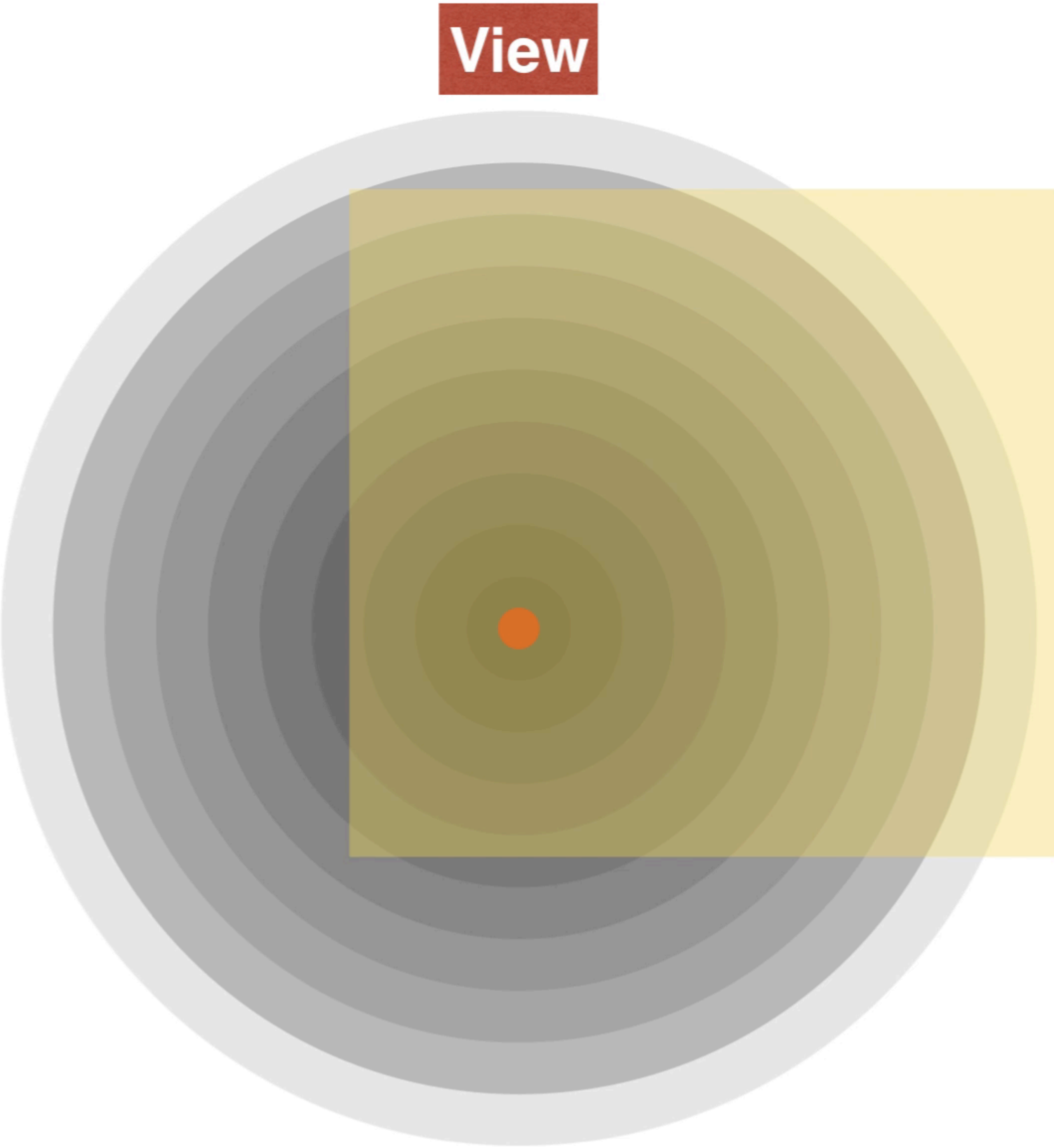
- Bayesian Information Gain (BIG) - BIGnav
  - **Maximizes the expected information gain** $IG(\Theta|X=x,Y)$

Go over all possible feedback, and find the one that maximizes the expected information gain

$$IG(\Theta|X=x,Y)=H(\Theta)-H(\Theta|X=x,Y)$$

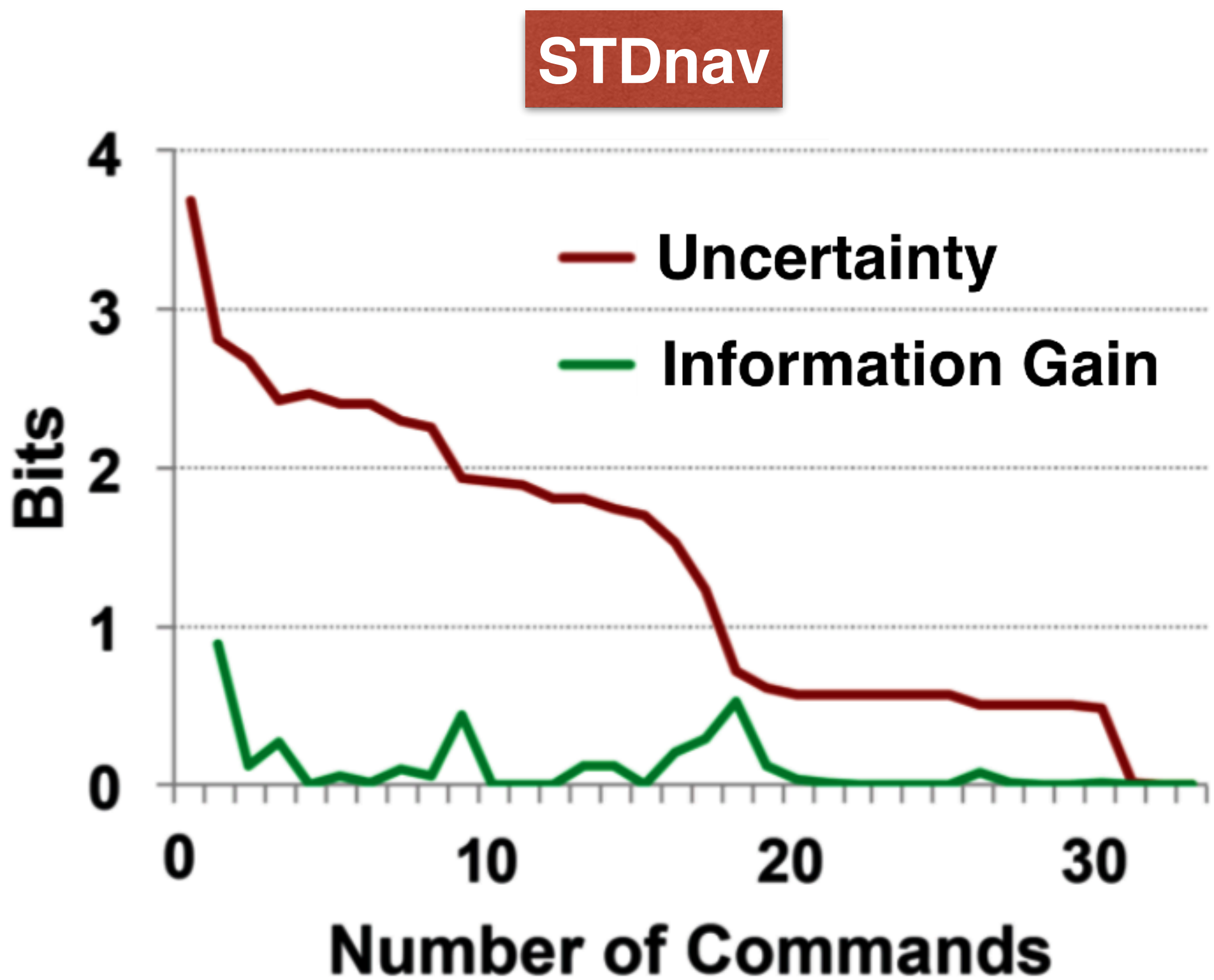


- Bayesian Information Gain (BIG) - BIGnav



- Bayesian Information Gain (BIG) - BIGnav

- **BIGnav gains maximum information from each user input**



A map application - “3 steps to go to Paris”.

Europe map featuring large cities with their population as distribution.

A map application - “Navigate to Helsinki”.

Europe map featuring large cities with their population as distribution.

Exercise

- Bayesian Information Gain (BIG) - BIGnav

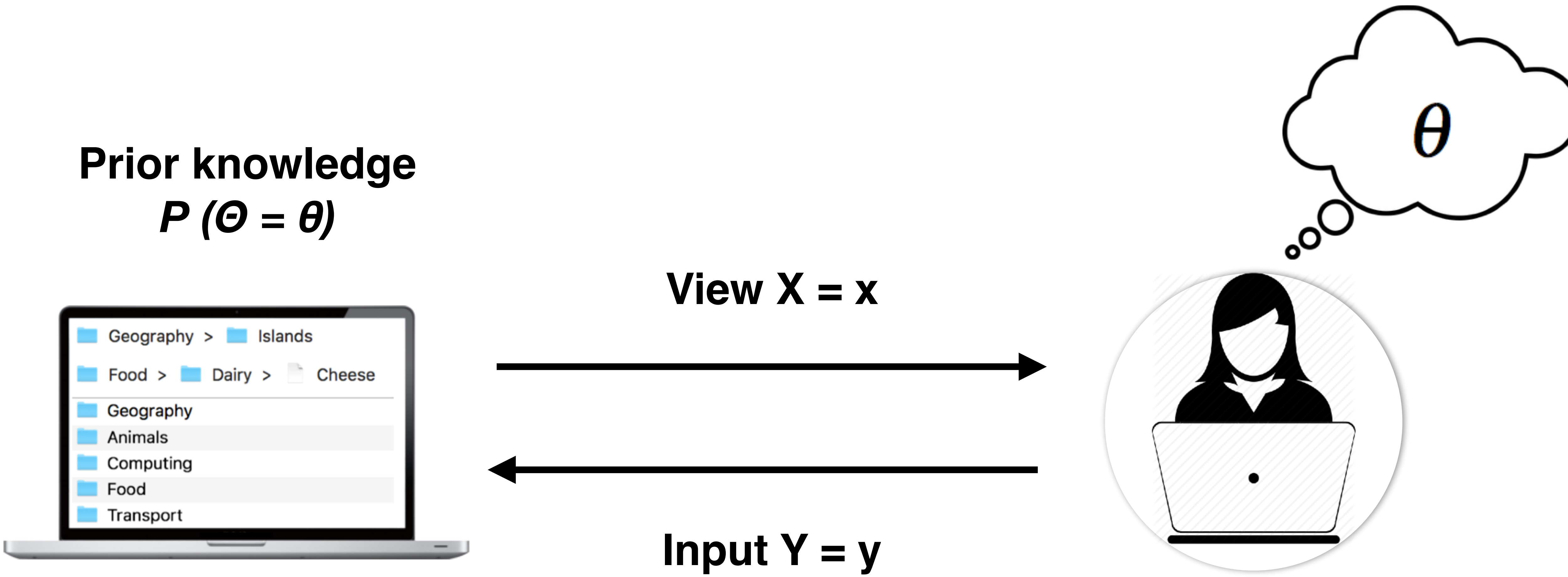
More efficient but…

Higher cognitive load

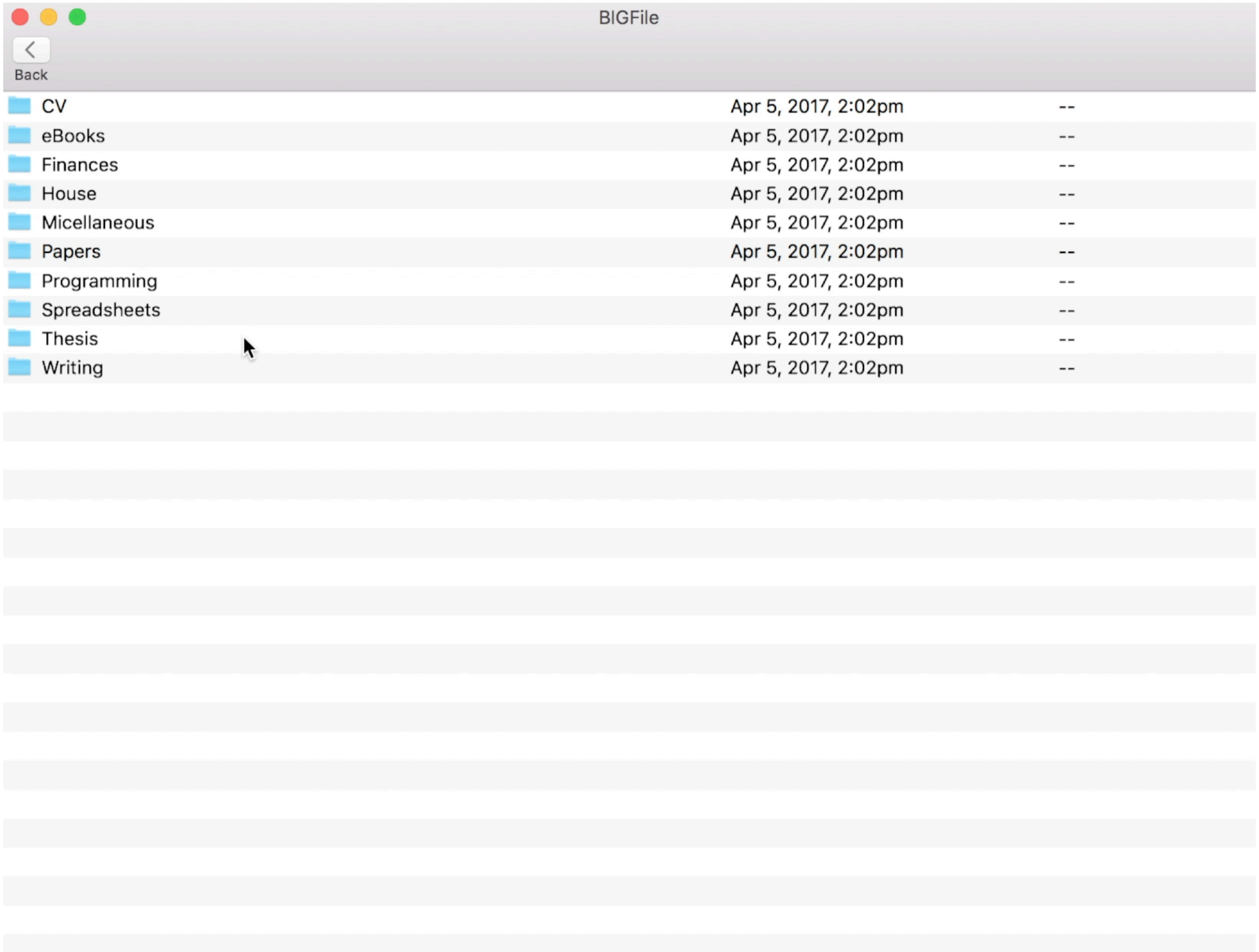
🙄

- Bayesian Information Gain (BIG) - BIGFile
    - **Leverages the expected information gain** $IG(\Theta|X=x,Y)$

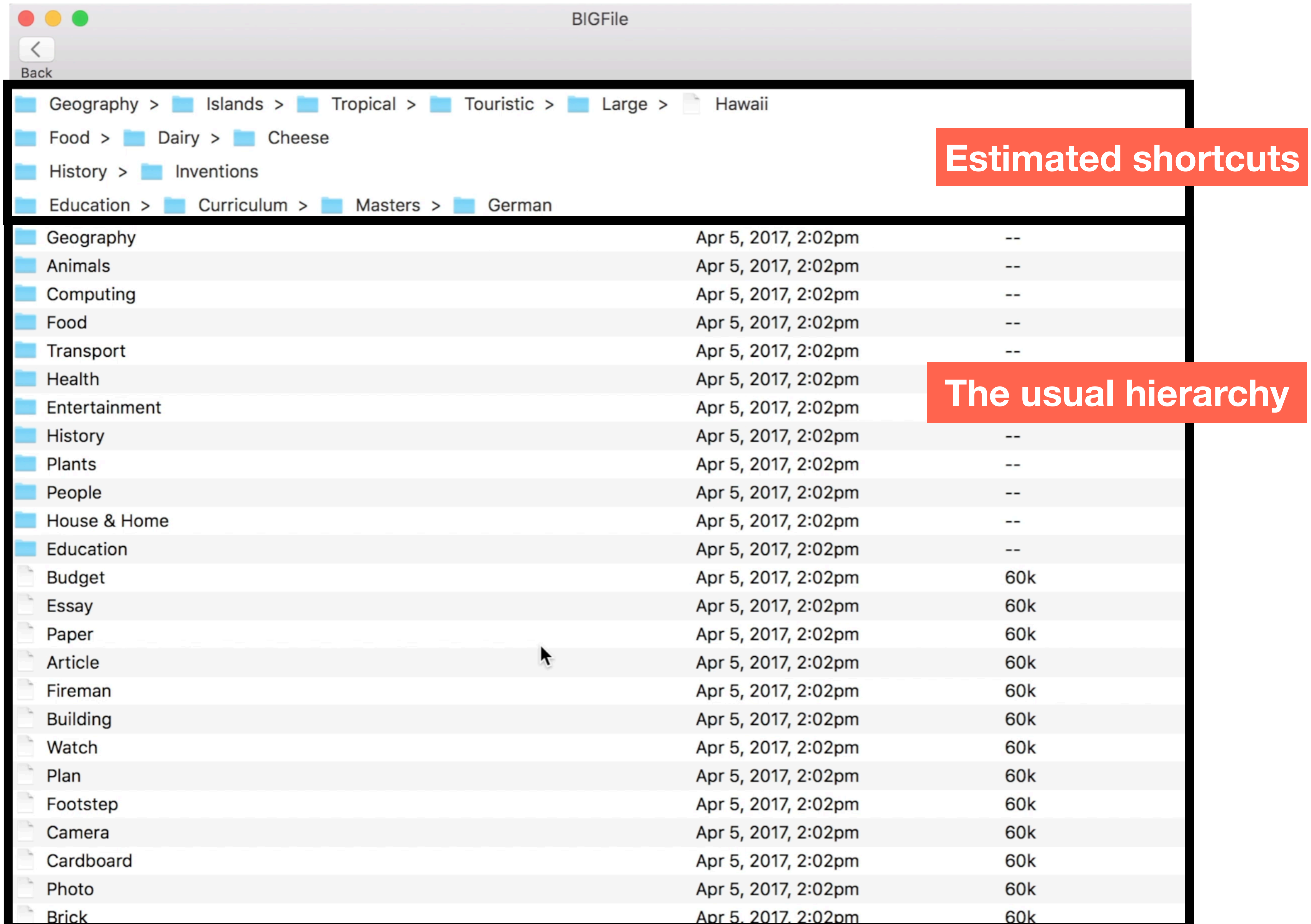
Prior knowledge
P (Θ = θ)

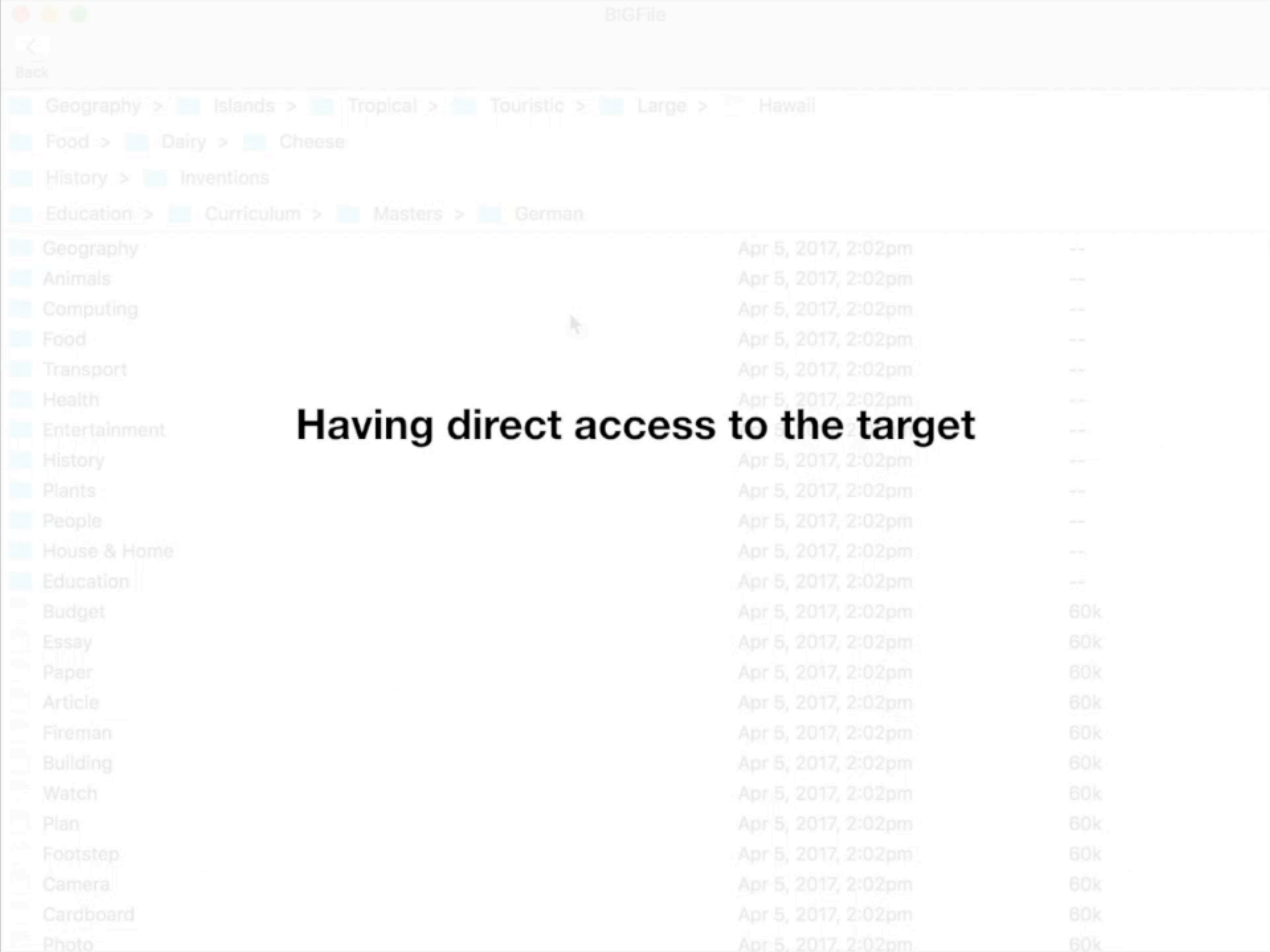


- Bayesian Information Gain (BIG) - BIGFile



- Bayesian Information Gain (BIG) - BIGFile





Having direct access to the target

- Bayesian Information Gain (BIG)

- Executes the user input only **Multiscale navigation**
- Maximizes the expected information gain $IG(\Theta|X=x,Y)$ **BIGnav**
- Leverages the expected information gain **BIGFile**

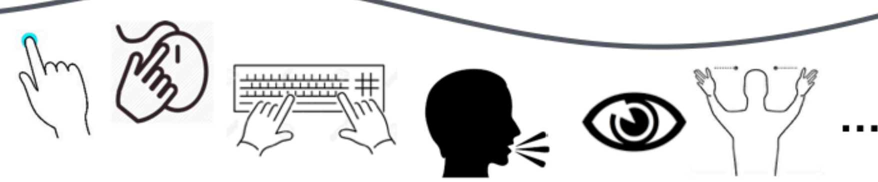
Feedback $X = x$

Computer

- **Prior knowledge:** $P(\Theta = \theta)$
- **User behavior model:** $P(Y = y|\Theta = \theta, X = x)$
- **Update knowledge:** $P(\Theta = \theta \mid X = x, Y = y)$
- **Calculate Information Gain:** $IG(\Theta|X = x, Y = y)$

User

Input $Y = y$

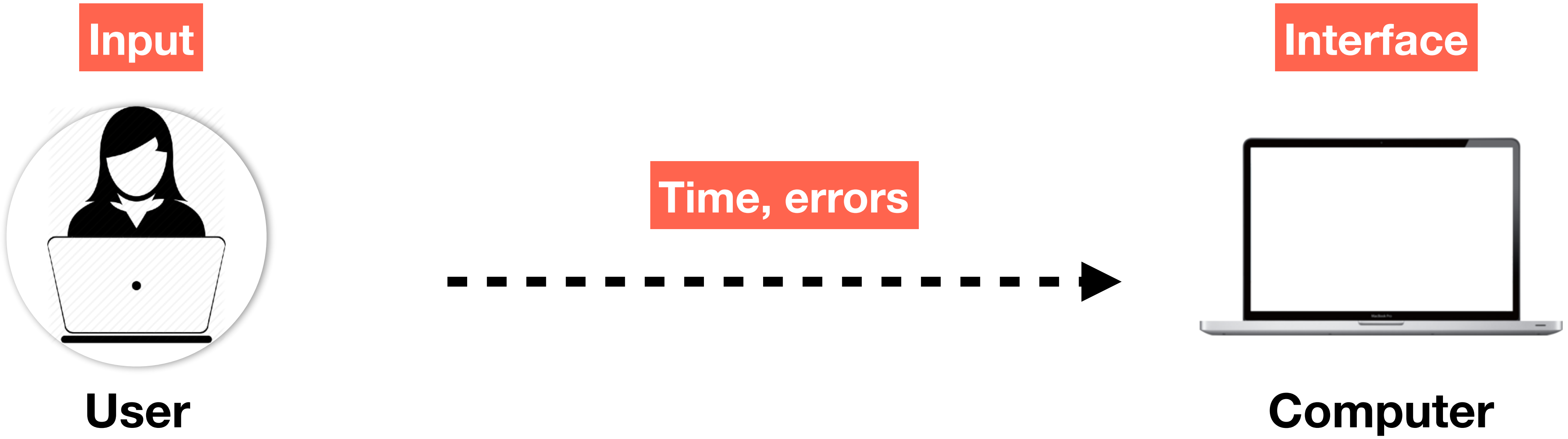


Coffee Break

🏆 **BIGnav:** Bayesian Information Gain for Guiding Multiscale Navigation. (CHI'17)

🎗 **BIGFile:** Bayesian Information Gain for Fast File Retrieval. (CHI'18)

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction

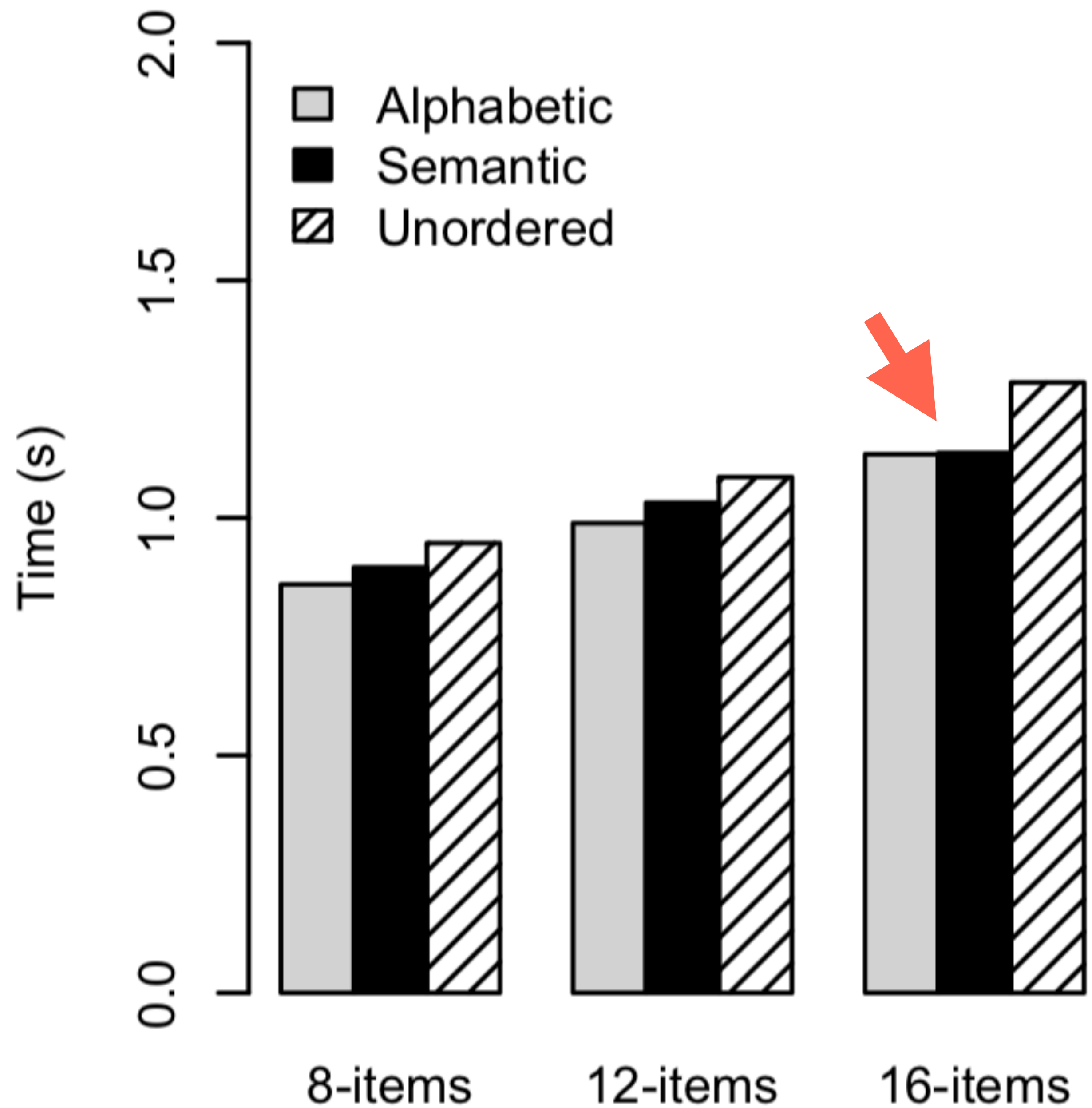
Drawback 1: Speed-accuracy tradeoff.

- Information-theoretic measures to characterize interaction

Solution : Control errors.

* Control error rate under 4 %.
* Remove errors from data analysis.

- Information-theoretic measures to characterize interaction

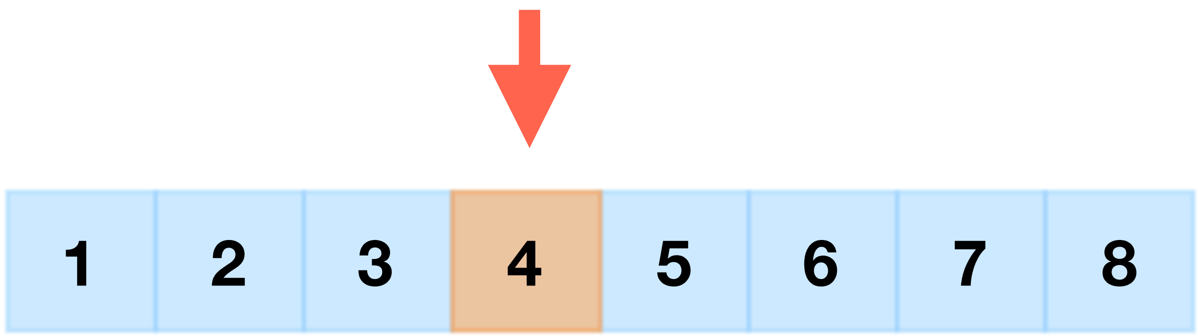


Model of visual search and selection time in linear menus. (Bailly et al. CHI'14)

- Information-theoretic measures to characterize interaction

Drawback 2: The treatment of errors.

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction

1	2	3	4	5	6	7	8

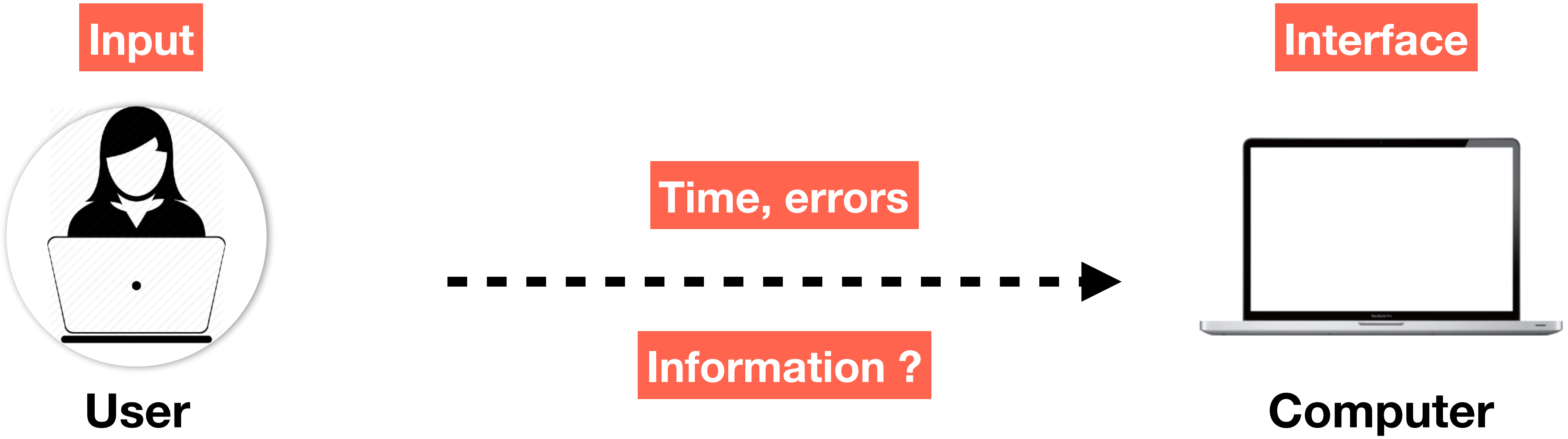
Jean-Pierre: {4, 4, 3, 4, 4, 5, 4, 3}

Error rate: 3 / 8 = 37.5 %

Chantal: {4, 4, 1, 4, 4, 8, 4, 7}

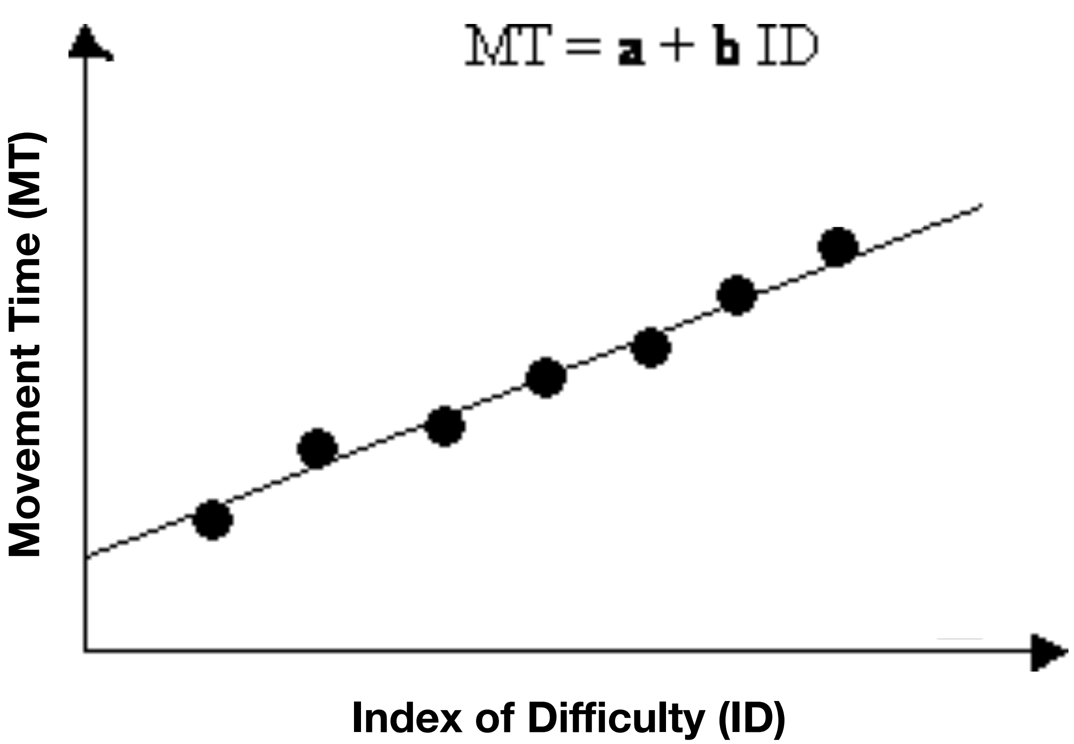
Error rate: 3 / 8 = 37.5 %

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction

Fitts' law

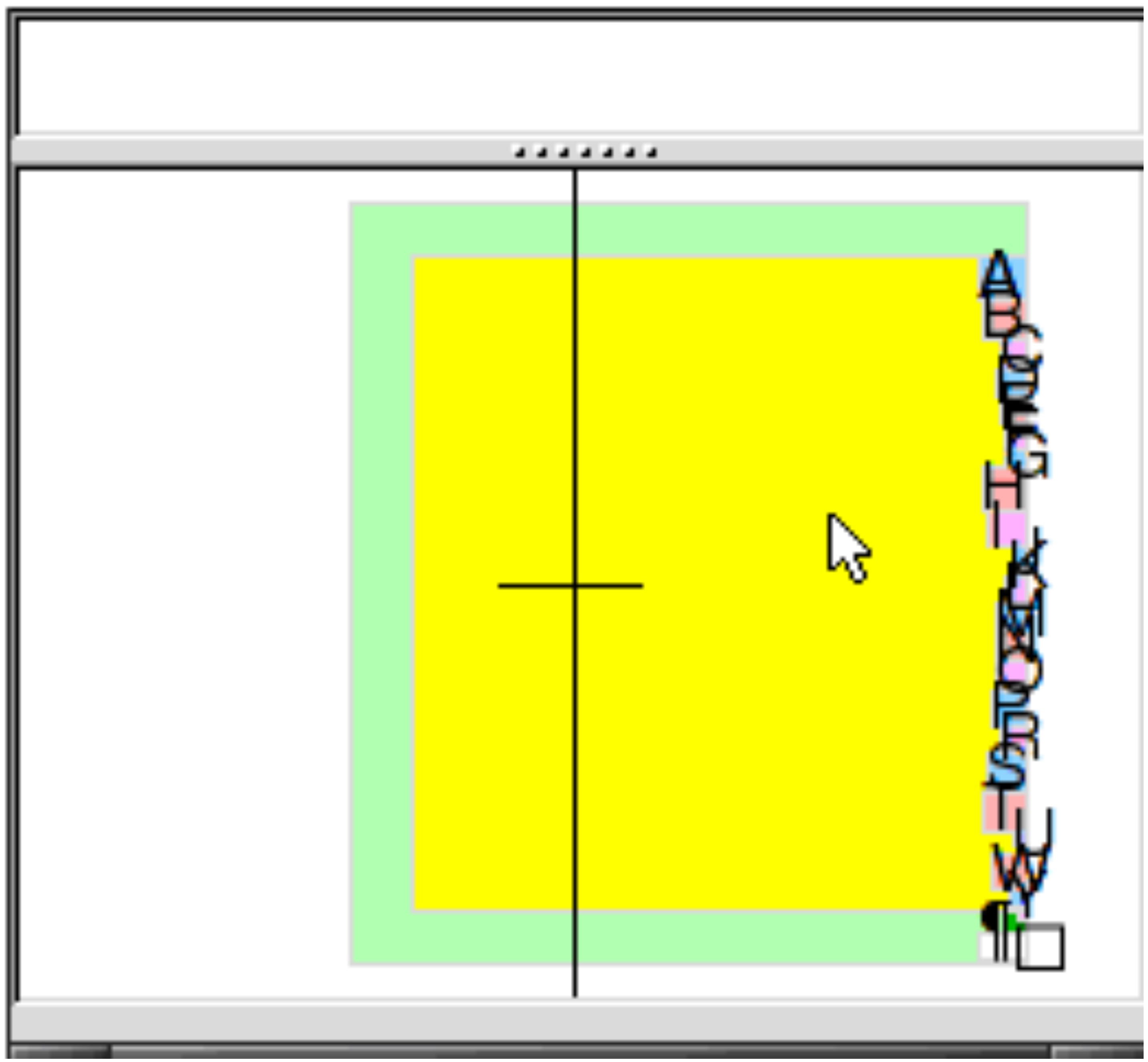


Movement Time: $T = a + b\log_2\left(\frac{A+W}{W}\right)$

Channel Capacity: $C = B\log_2\left(\frac{S+N}{N}\right)$

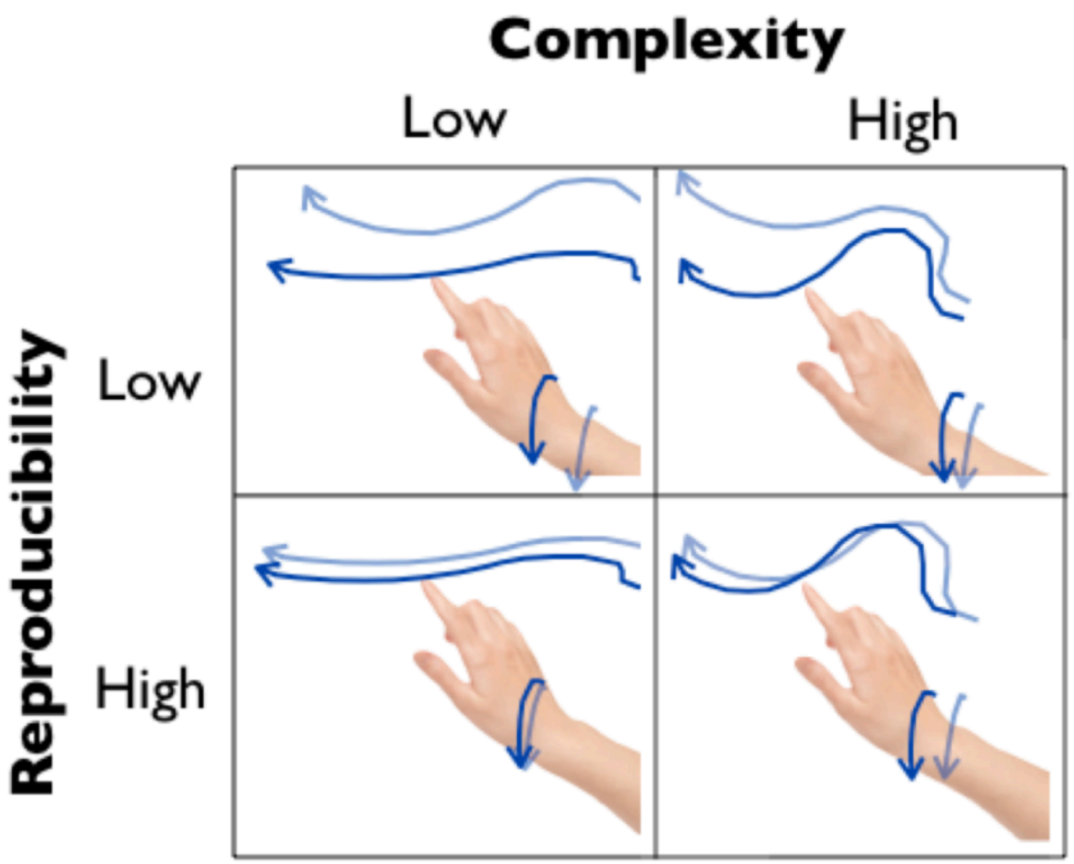
- Information-theoretic measures to characterize interaction

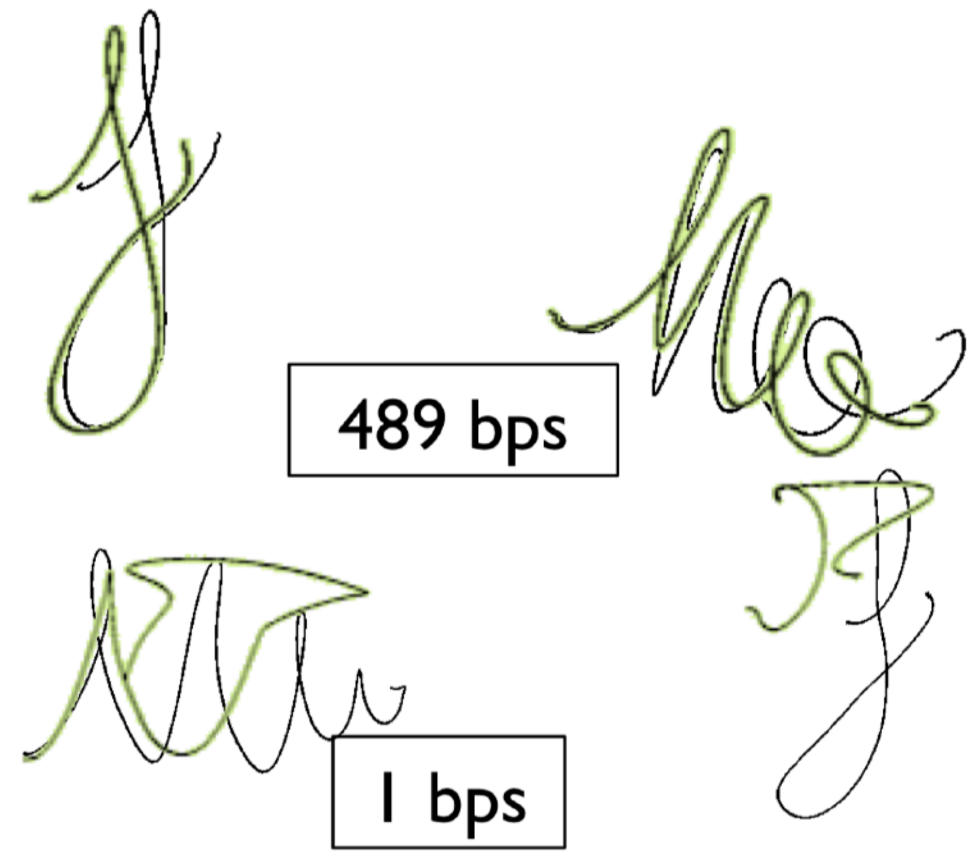
Dasher



Source: http://www.inference.org.uk/dasher/

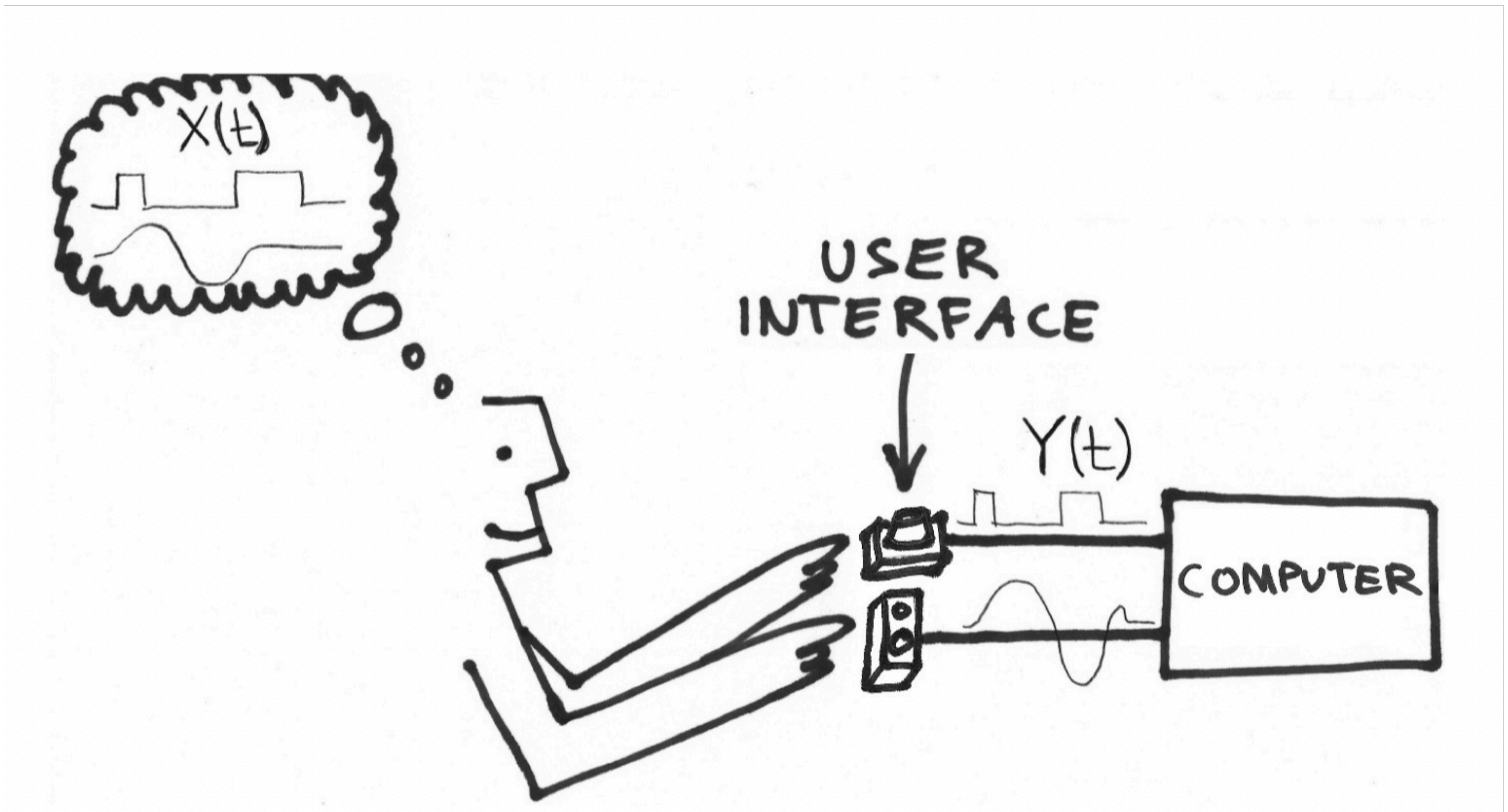
- Information-theoretic measures to characterize interaction





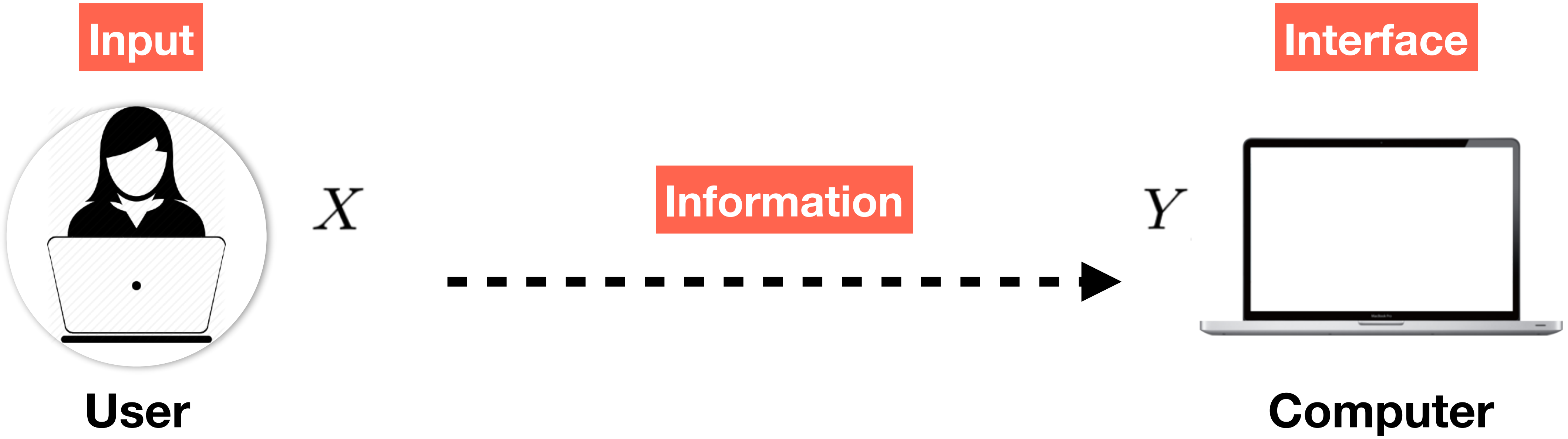
Information Capacity of Full-body Movements. (Oulasvirta et al. CHI'13)

- Information-theoretic measures to characterize interaction

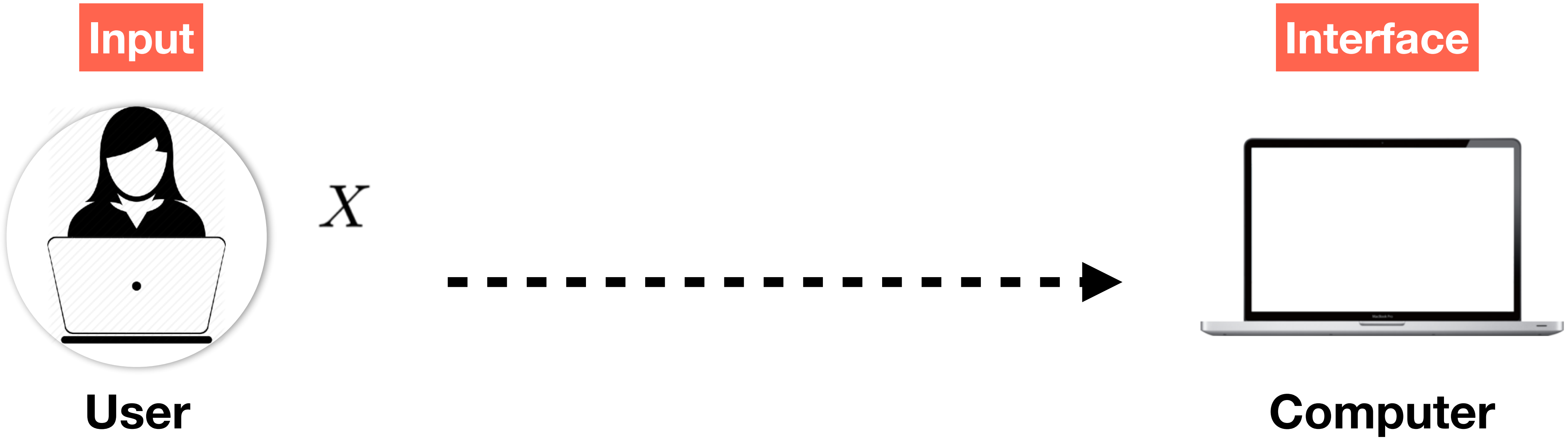


An Approach for Using Information Theory to Investigate Continuous Control of Analog Sensors by Humans. (Berdahl et al. AM'16)

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction



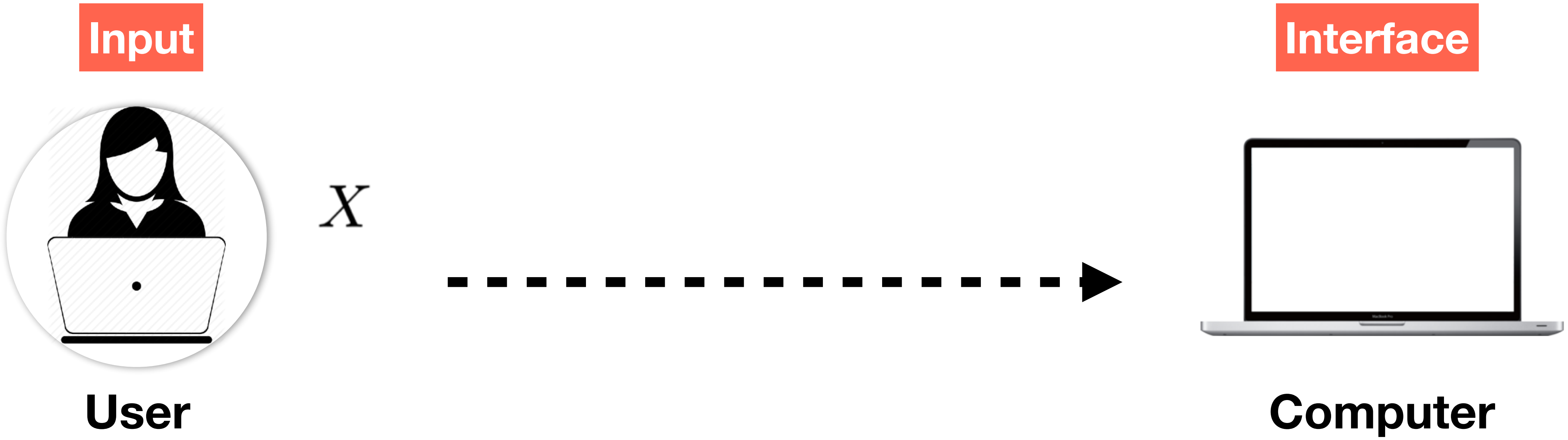
X: A set of all possible messages that a user can transmit, representing the intended inputs.

x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

x_1 x_2 x_3 x_4

X takes values in {Copy, Paste, Help, Share}

- Information-theoretic measures to characterize interaction



P(X): The probability distribution of the intended inputs.

$P(x_1)$ $P(x_2)$ $P(x_3)$ $P(x_4)$

X takes values in {Copy, Paste, Help, Share}

x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

- Information-theoretic measures to characterize interaction

H(X): Input entropy: How much information could be transmitted. Corresponding to ***input size*** and the ***probability distribution***.

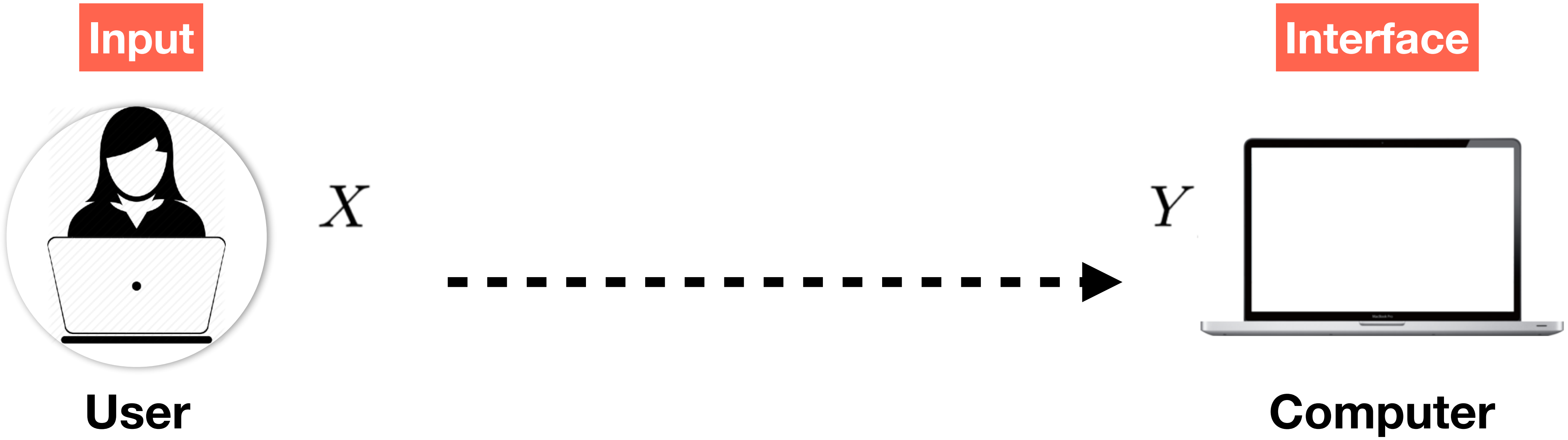
x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

X takes values in {Copy, Paste, Help, Share}

$$H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

$$0 \le H(X) \le \log N$$

- Information-theoretic measures to characterize interaction



Y: The actual input received by the computer.

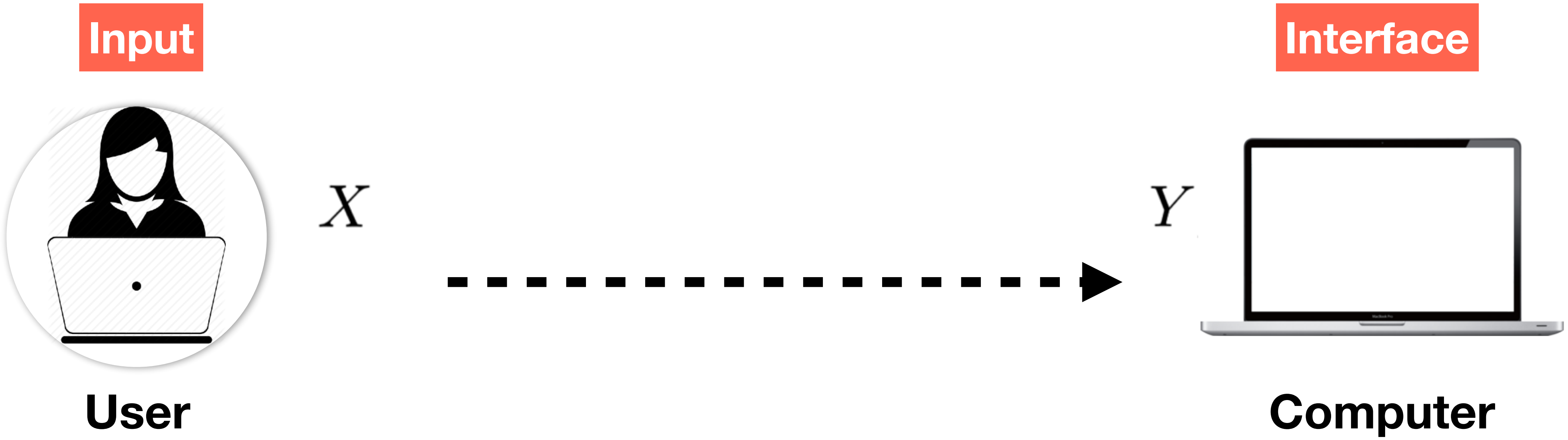
x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

X = [Copy, Paste, Copy, Paste, Help, Share, Help, Share]

H(X) = 2 bits

Y = [Copy, Paste, Copy, Paste, *Share*, Share, Help, Share]

- Information-theoretic measures to characterize interaction



I(X;Y): Mutual information between the intended input and the actual input. It describes how much information actually gets transmitted.

x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

$$I(X;Y) = \sum_x \sum_y P(X=x, Y=y) \log_2 \frac{P(X=x, Y=y)}{P(X=x)P(Y=y)}$$

$$I(X;Y) = H(X) - H(X|Y)$$

- Information-theoretic measures to characterize interaction

I(X;Y): Mutual information between the intended input and the actual input. It describes how much information actually gets transmitted.

x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

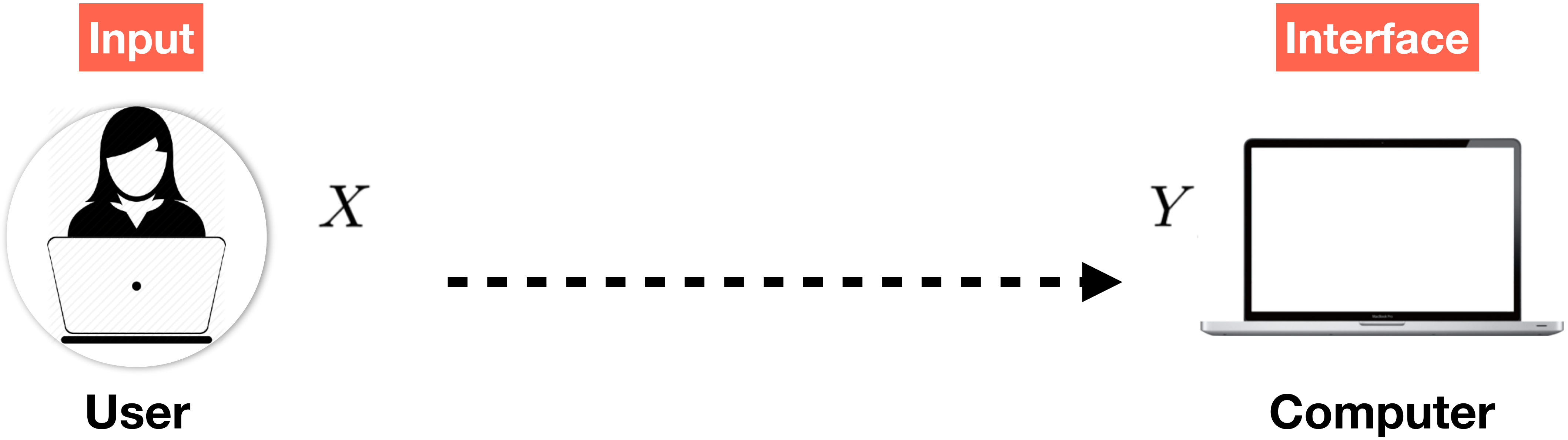
X = [Copy, Paste, Copy, Paste, Help, Share, Help, Share]

H(X) = 2 bits

Y = [Copy, Paste, Copy, Paste, *Share*, Share, Help, Share]

I(X;Y) = 1.656 bits

- Information-theoretic measures to characterize interaction



PI = I(X;Y) / H(X): Information transmission percentage.

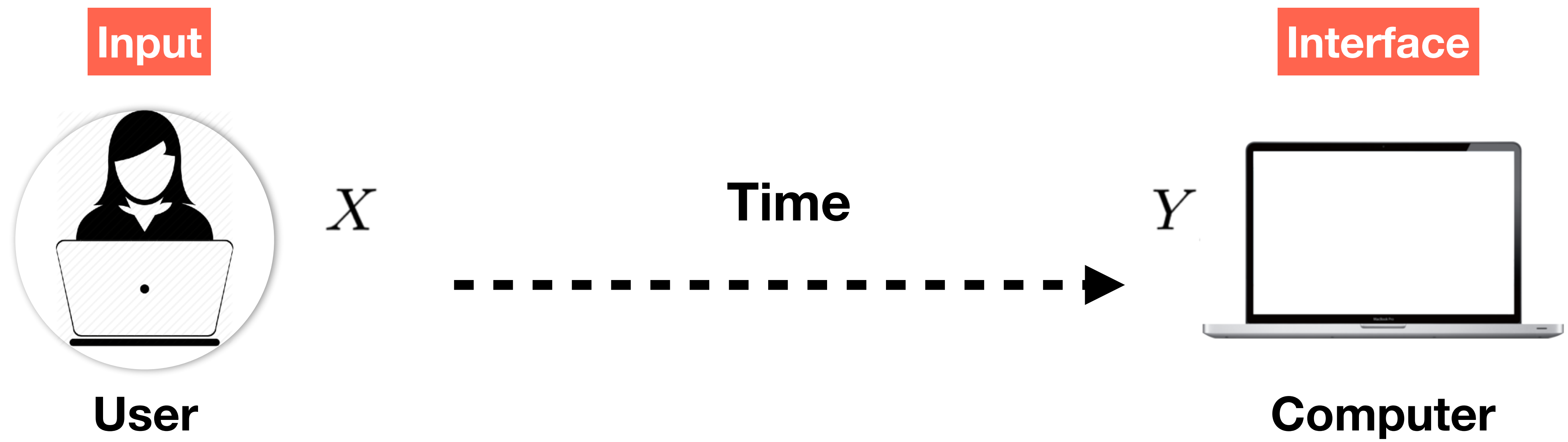
x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

H(X) = 2 bits

I(X;Y) = 1.656 bits

PI = I(X;Y) / H(X) = 1.656 / 2 = 82.8%

- Information-theoretic measures to characterize interaction



TP = I(X;Y) / T: Throughput - information transmission efficiency.

x_1	**Copy**
x_2	**Paste**
x_3	**Help**
x_4	**Share**

H(X) = 2 bits

I(X;Y) = 1.656 bits

T = 1.5 s

TP = I (X;Y) / T = 1.656 / 1.5 = 1.104 bit/s

- Information-theoretic measures to characterize interaction

Advantage 1: Standard language to describe interaction.

- Information-theoretic measures to characterize interaction

Advantage 1: Standard language to describe interaction.

H(X): how much information could be transmitted.

I(X;Y): how much information actually gets transmitted.

H(X;Y): how much information gets lost, related to how users make errors.

TP: information transmission efficiency.

- Information-theoretic measures to characterize interaction

Advantage 2: Theoretical foundation.

- Information-theoretic measures to characterize interaction

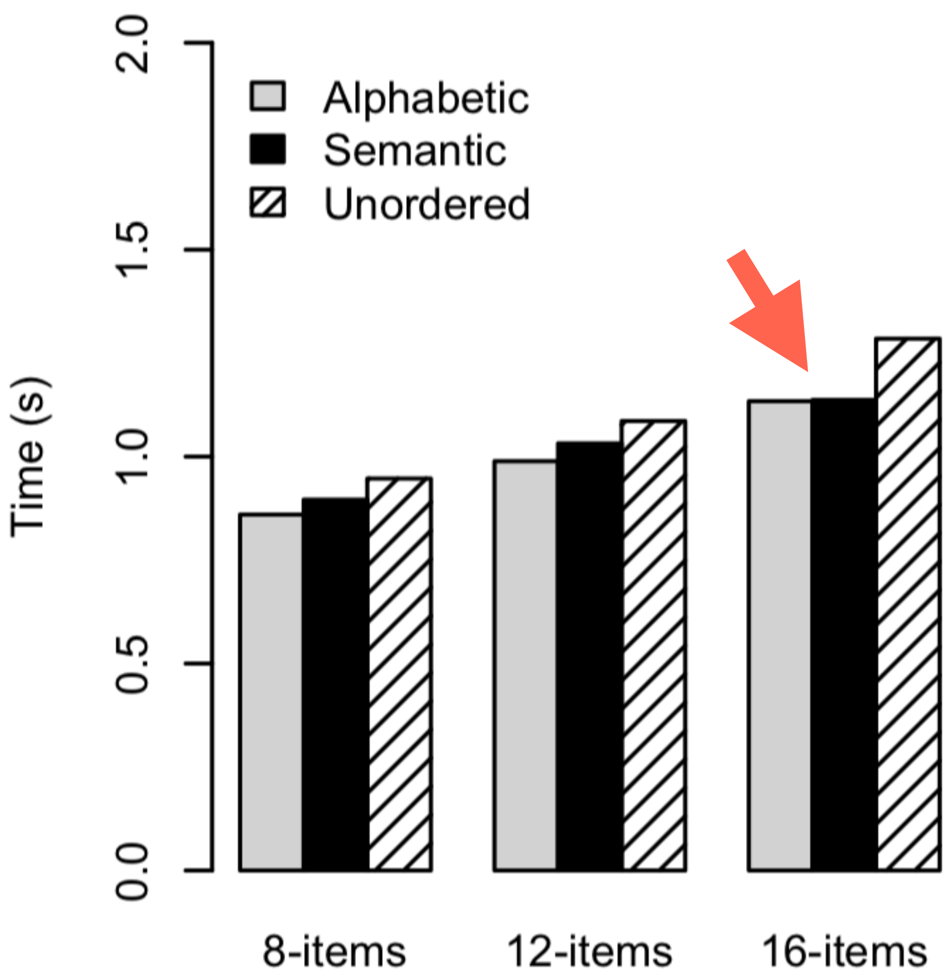
Advantage 2: Theoretical foundation.

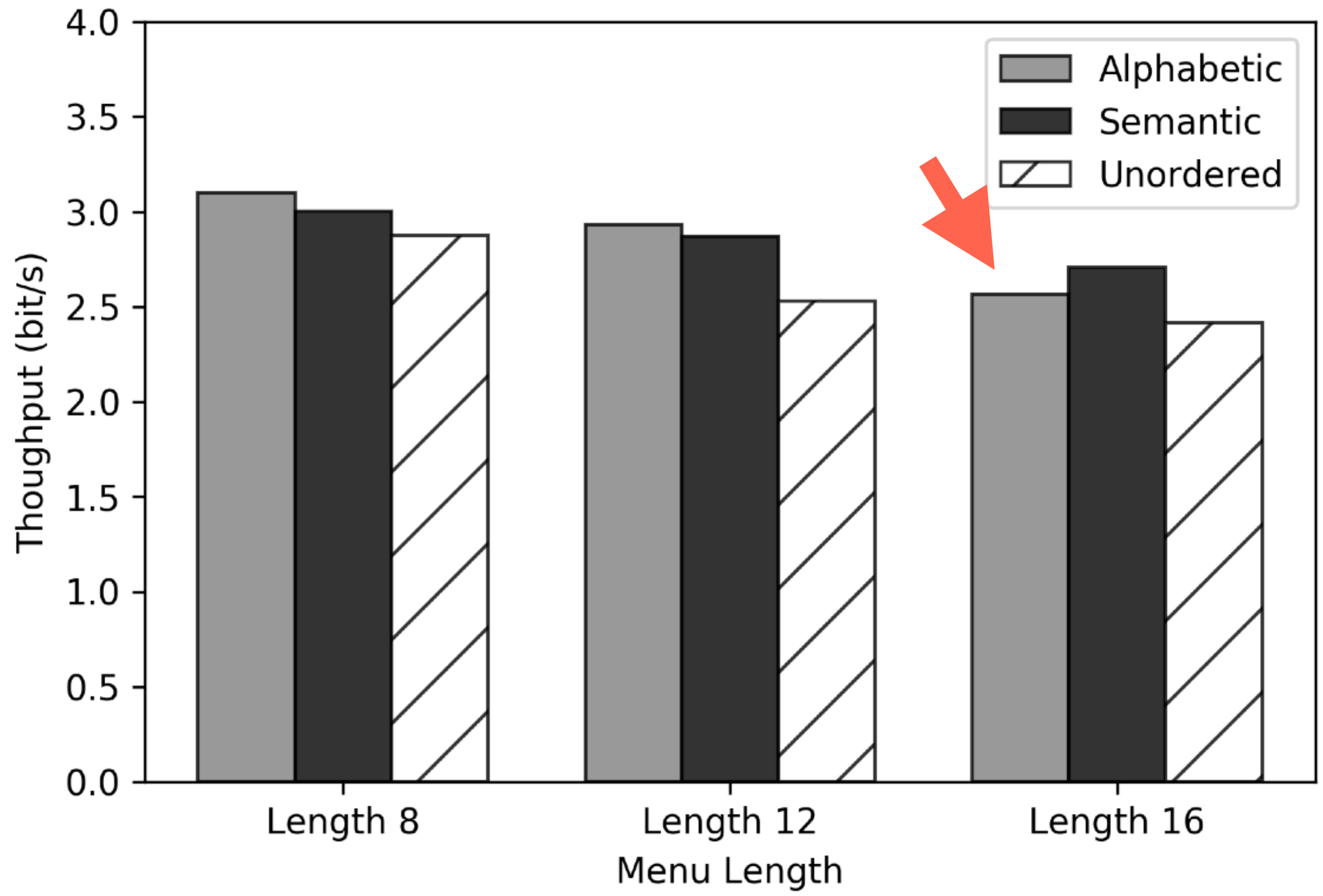
	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP
NofX $\uparrow$	$\uparrow$	$\uparrow$	-	$\uparrow$
$p(X)$ $\uparrow$	$\downarrow$	$\downarrow$	-	$\downarrow$
P_e $\uparrow$	-	$\downarrow$	$\uparrow$	$\downarrow$
T $\uparrow$	-	-	-	$\downarrow$

- Information-theoretic measures to characterize interaction

Advantage 3: Speed-accuracy tradeoff.

- Information-theoretic measures to characterize interaction



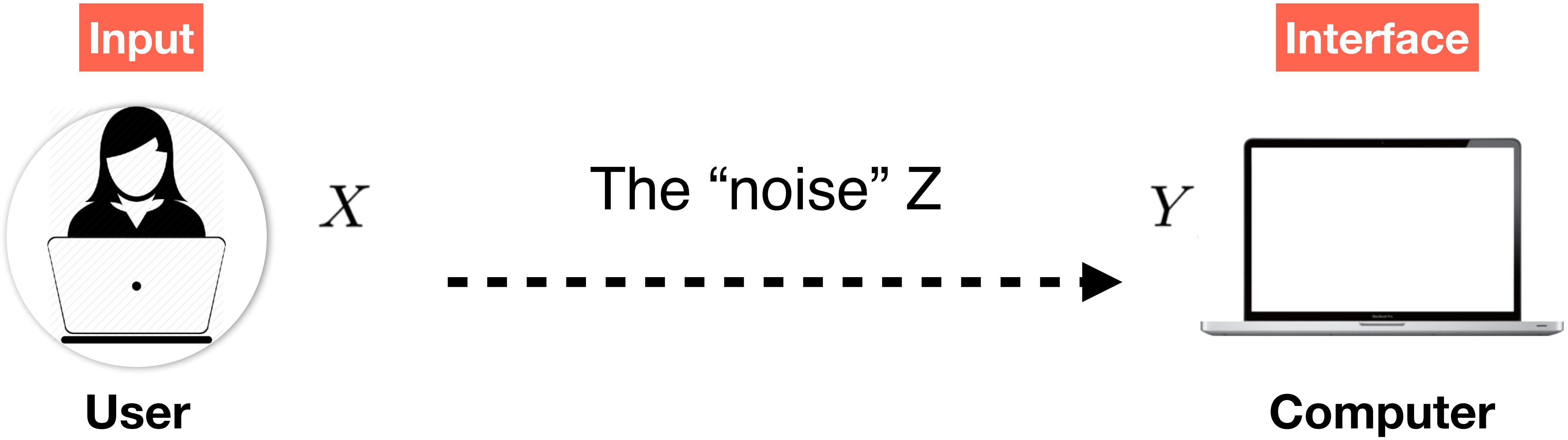


Model of visual search and selection time in linear menus. (Bailly et al. CHI'14)

- Information-theoretic measures to characterize interaction

Advantage 4: Equivocation provides information about how users make errors.

- Information-theoretic measures to characterize interaction



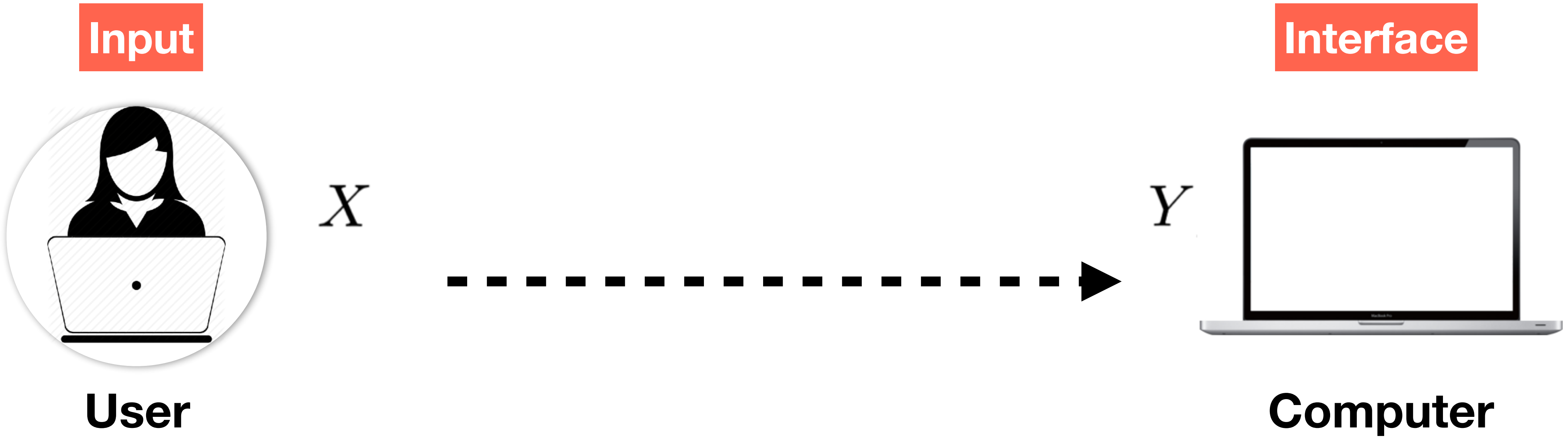
The error random variable:

$$E = \begin{cases} 0 & \text{if } X = Y; \\ 1 & \text{if } X \neq Y. \end{cases}$$

The error rate:

$$P_e = P(X \neq Y)$$

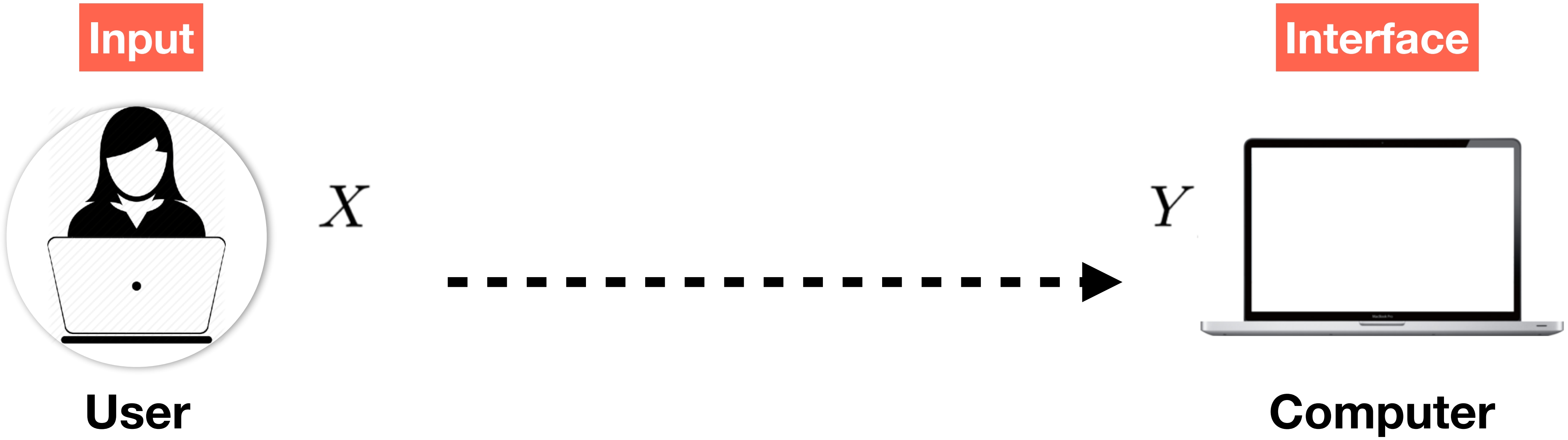
- Information-theoretic measures to characterize interaction



The error rate P_e has binary entropy:

$$H(E) = -P_e \log_2 P_e - (1 - P_e) \log_2 (1 - P_e)$$

- Information-theoretic measures to characterize interaction



$$I(X;Y) = H(X) - H(X|Y)$$

$$H(X|Y) \leq H(E) + P_e \times H(Z|E=1)$$

Fano's inequality.

[Theorem 2.10.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).

- Information-theoretic measures to characterize interaction

$$I(X;Y) = H(X) - H(X|Y)$$

$$H(X|Y) \leq H(E) + P_e \times H(Z|E=1)$$

The fact that users make errors. At most 1 bit.

- Information-theoretic measures to characterize interaction

$$I(X;Y) = H(X) - H(X|Y)$$

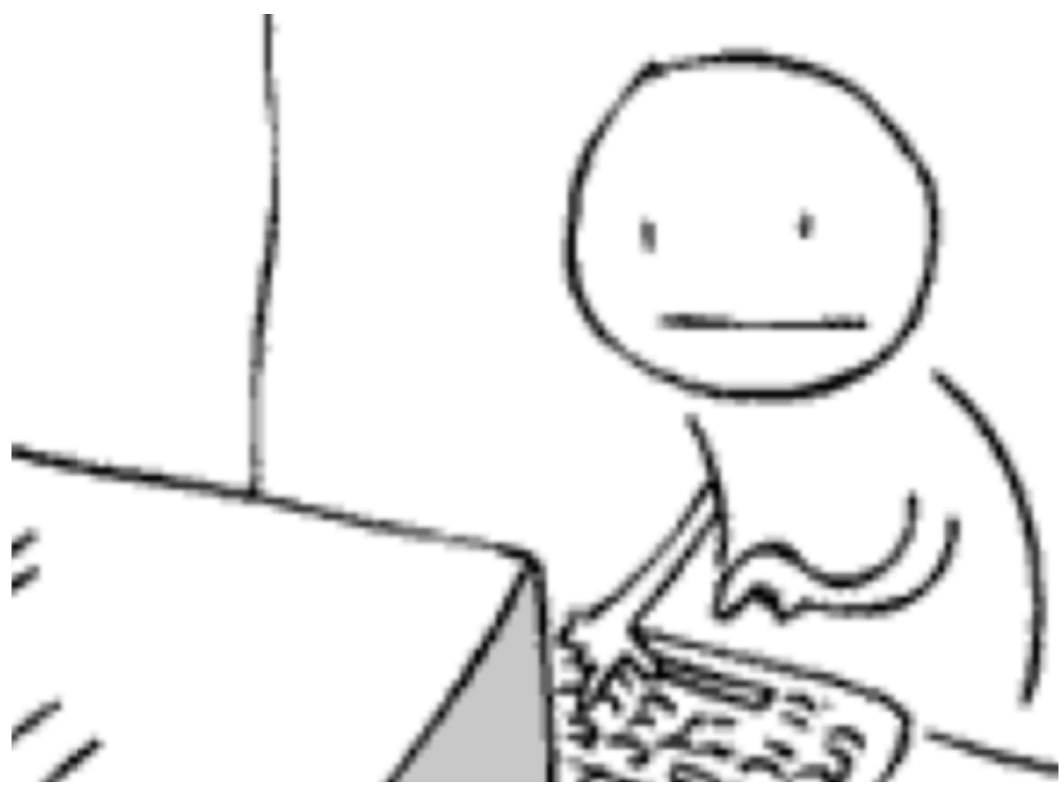
$$H(X|Y) \leq H(E) + \boxed{P_e \times H(Z|E=1)}$$

How they make errors.

Exercise

- Information-theoretic measures to characterize interaction

From independent variables to dependent variables…



- Information-theoretic measures to characterize interaction

X_2

Chantal: Let’s have Fish&Chips.

X_1 $\dots$ X_m

$$H(X_1, X_2, \dots, X_m)$$

N-gram language model:

$$p(X_m | X_{m-1}, X_{m-2}, \dots, X_{m-n})$$

- Information-theoretic measures to characterize interaction

1. Zero-order approximation (symbols independent and equiprobable).

   XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

2. First-order approximation (symbols independent but with frequencies of English text).

   OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (digram structure as in English).

   ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

A mathematical theory of communication (C.E.Shannon, 1948)

- Information-theoretic measures to characterize interaction

4. Third-order approximation (trigram structure as in English).

   IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

5. First-order word approximation. Rather than continue with tetragram, . . . , *n*-gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

   REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

   THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

A mathematical theory of communication (C.E.Shannon, 1948)

- Information-theoretic measures to characterize interaction

X = “ASK L”

Uni-gram model.

- Information-theoretic measures to characterize interaction

$$H(X_1, X_2, X_3, X_4, X_5) = \sum_{i=1}^{5} H(X_i | X_{i-1})$$

$$= H(X_1) + H(X_2|X_1) + H(X_3|X_2) + H(X_4|X_3) + H(X_5|X_4)$$

[Theorem 2.2.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).

- Information-theoretic measures to characterize interaction

Y = “ASK L”

[Theorem 2.2.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).

- Information-theoretic measures to characterize interaction

$$H(X_1, ..., X_m | Y_1, ... Y_m) = \sum_{i=1}^{m} H(X_i | X_{i-1}, Y_i, Y_{i-1})$$

$$= H(X_1|Y_1) + H(X_2|X_1, Y_1, Y_2) + ... + H(X_m|X_{m-1}, Y_{m-1}, Y_m)$$

[Theorem 2.2.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).

- Information-theoretic measures to characterize interaction

$$
\begin{aligned}
&I(X_1, X_2, \ldots, X_m | Y_1, Y_2, \ldots, Y_m) \\
&= H(X_1, X_2, \ldots, X_m) - H(X_1, X_2, \ldots, X_m | Y_1, Y_2, \ldots, Y_m)
\end{aligned}
$$

[Theorem 2.2.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).

Try it yourself

Group work

* Implement the intelligent text entry method that you learned earlier
* Try a good & a bad decoder to see how the measures change
* Discuss the benefits of using information measures and how to use them for design

Email: wanyu.liu@telecom-paristech.fr
Web: http://perso.telecom-paristech.fr/wliu/