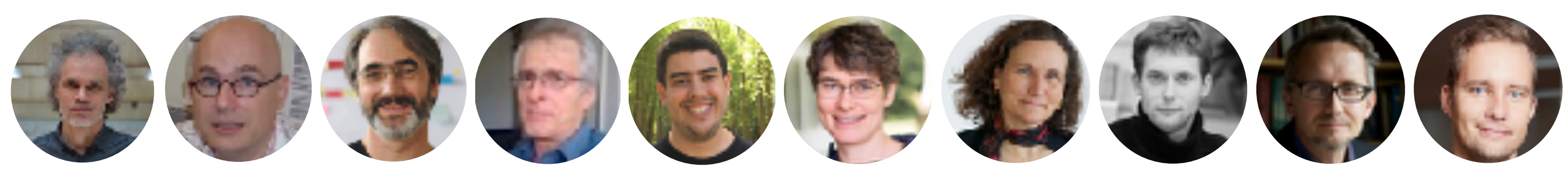
A Bayesian Information Gain (BIG) Approach for HCI Applications

Abby Wanyu Liu

IRCAM Centre Pompidou

About Me 

* Postdoc at IRCAM Centre Pompidou, Paris

* Working with 

* Keywords: HCI, information theory, computational interaction, bayesian, movement sonification

5th Summer School on Computational Interaction

Inference, optimization and modeling for the engineering of interactive systems

July 29th - August 2nd, 2019

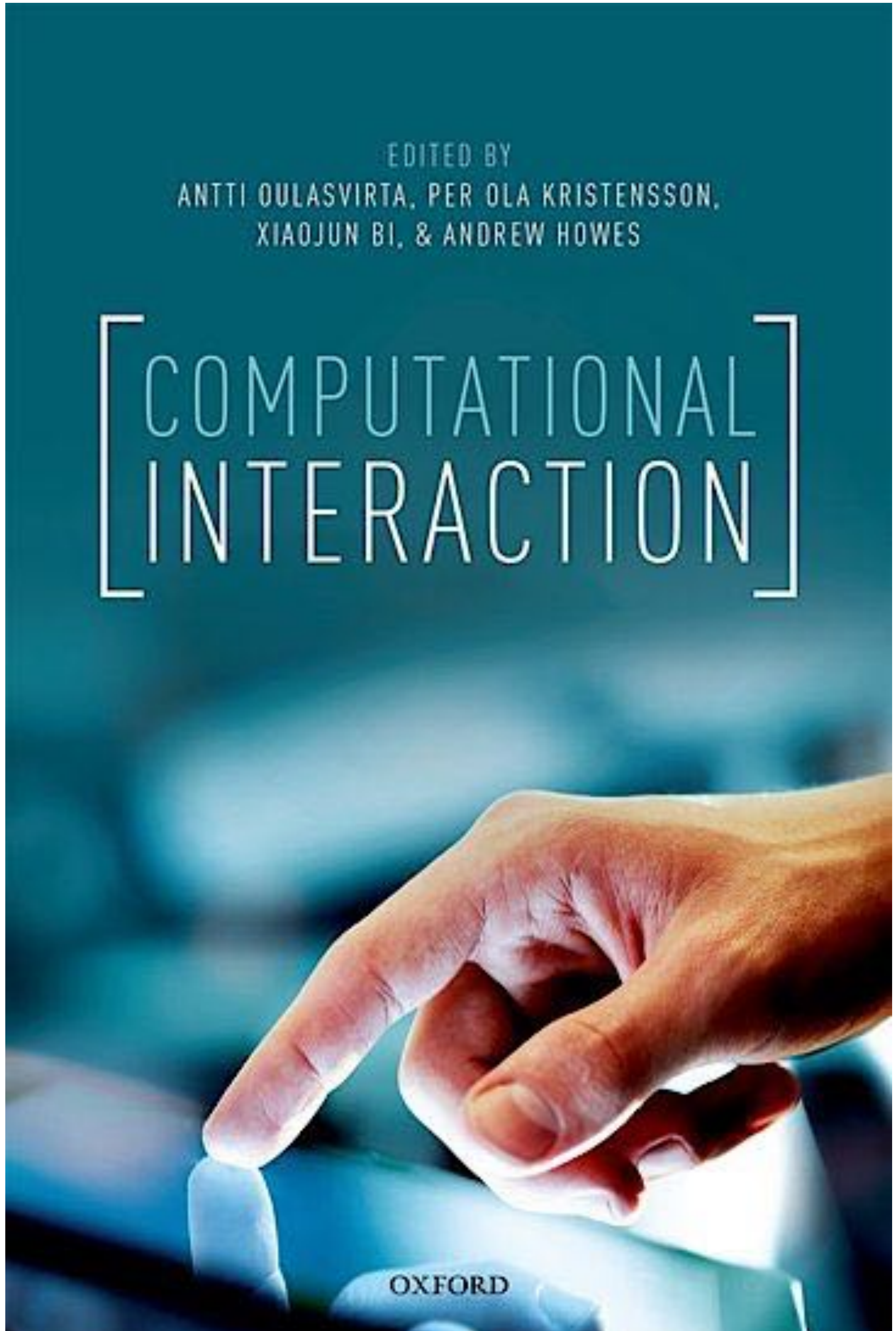
Columbia University and Stony Brook University, USA


2015: University of Glasgow


2016: Aalto University, Helsinki


2017: ETH Zurich


2018: University of Cambridge


EDITED BY
ANTTI OULASVIRTA, PER OLA KRISTENSSON,
XIAOJUN BI, & ANDREW HOWES
COMPUTATIONAL
INTERACTION
OXFORD

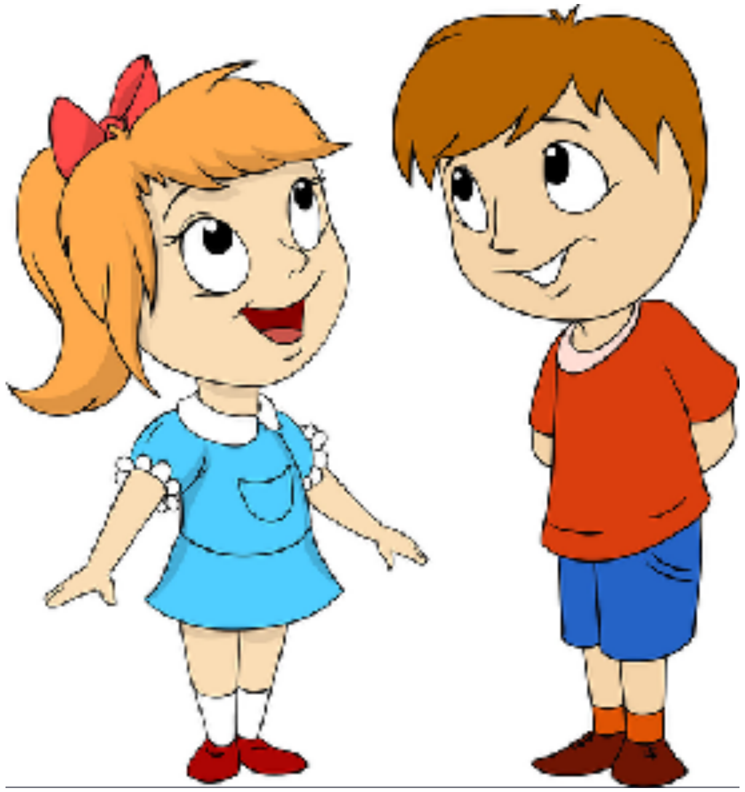
Agenda

* 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/5thComputationalInteraction

- Information is everywhere

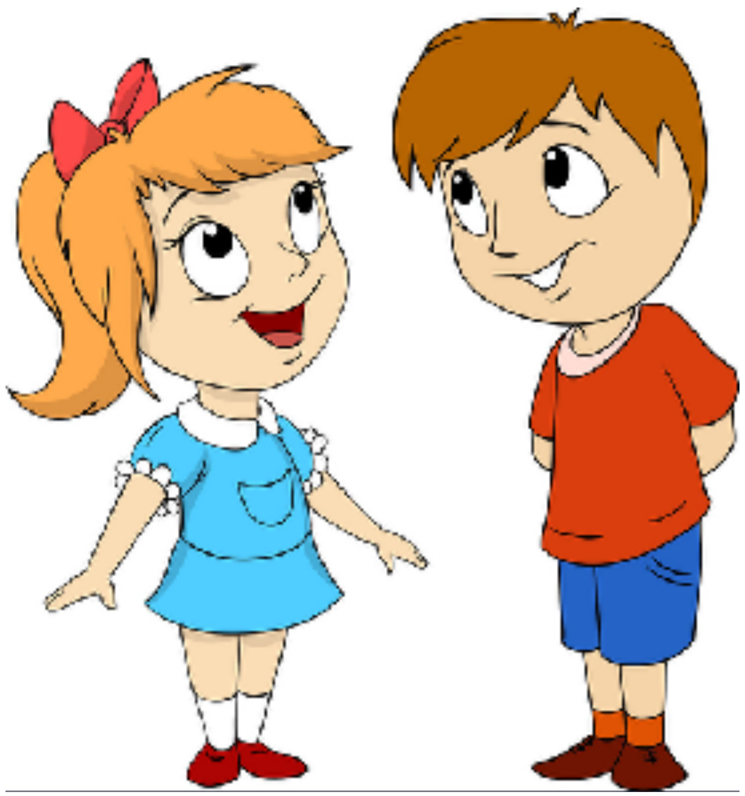


Jean-Pierre: Where do we go for dinner tonight? Pizza or burger?

Chantal: Let’s have pizza!

Information

- Information is everywhere



Jean-Pierre: Where do we go for dinner tonight? Pizza or burger?

Chantal: I don't know. We had pizza the other day so I would feel more like burger tonight. But I really like pizza. 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 !

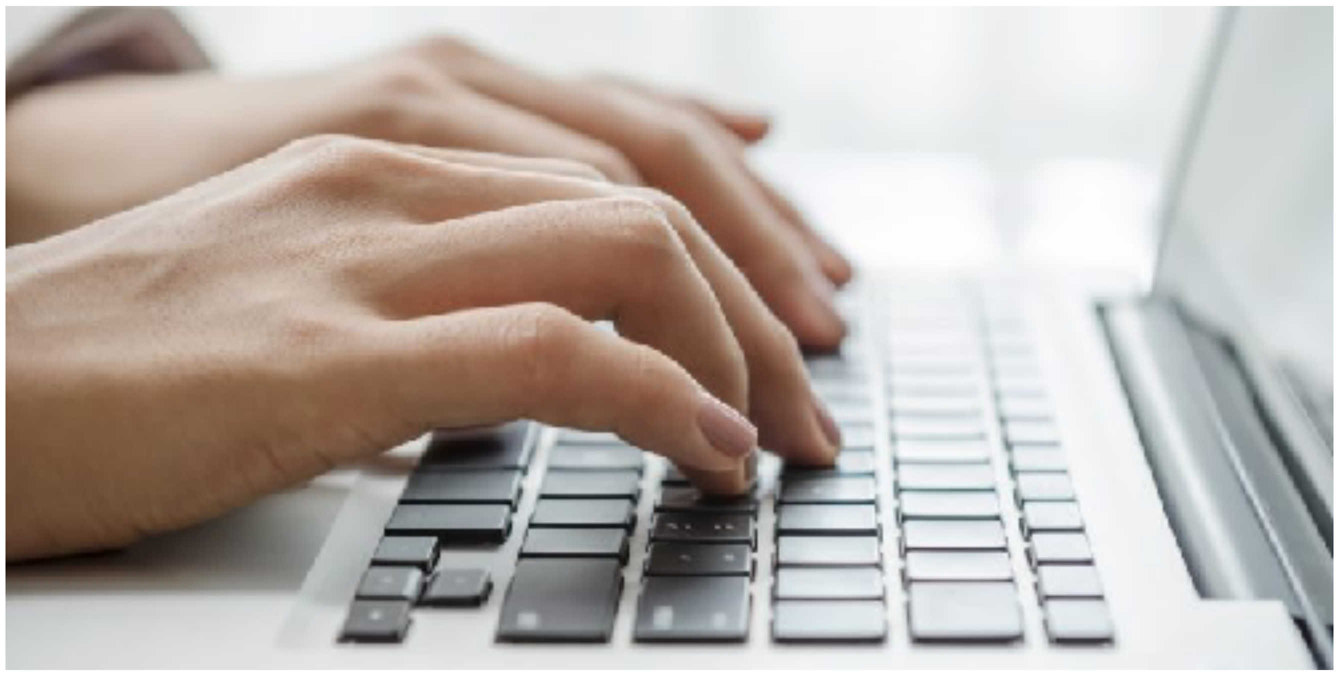
- Information is everywhere

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

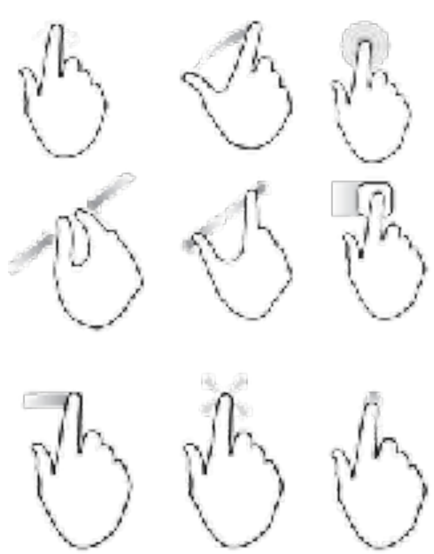
- Information in human-computer interaction

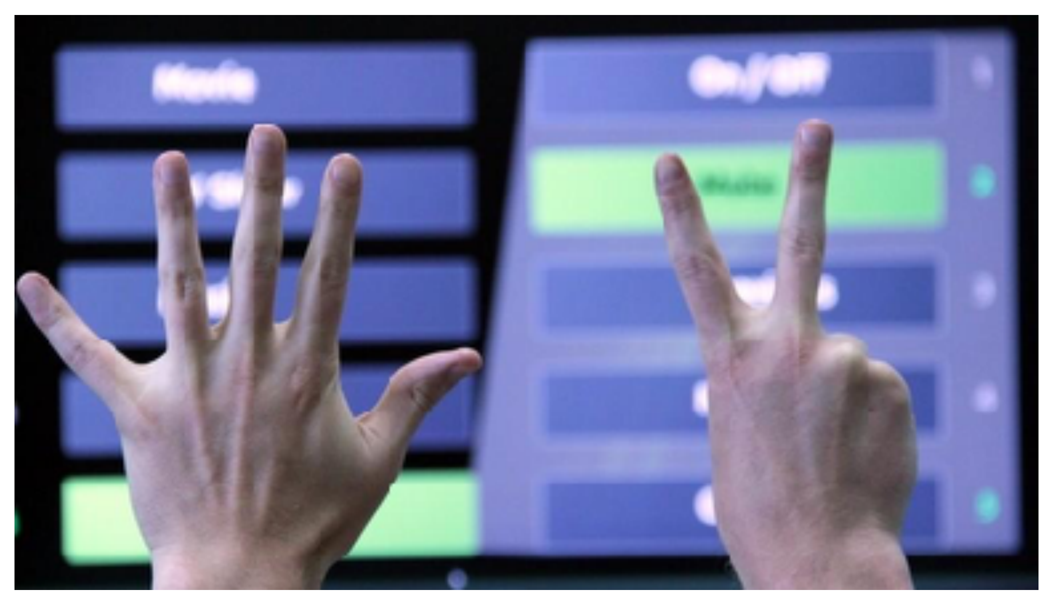
Characters



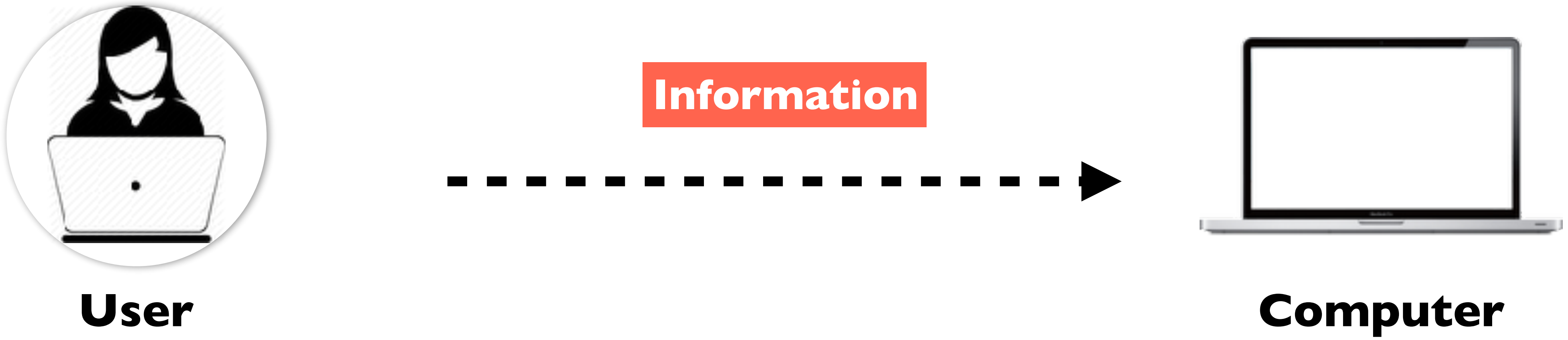
Words

Gestures





- Information in human-computer interaction

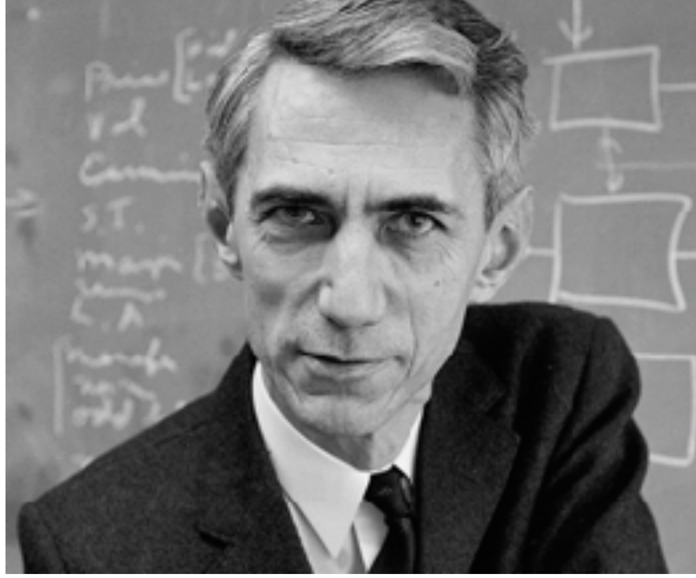


Can we quantify information?

Information in a more general sense?

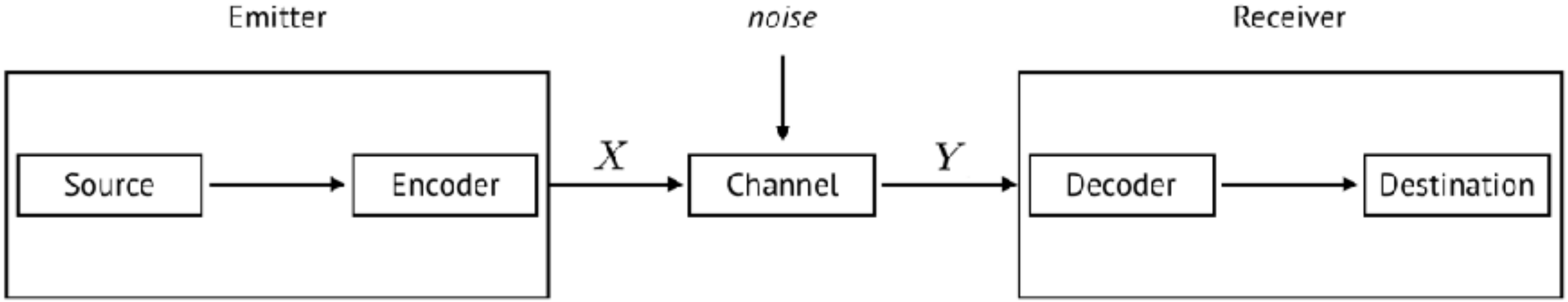
HCI as information transmission in *What Is Interaction?* (Kasper Hornbæk & Antti Oulasvirta, CHI'17)

- 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

Mathematics, statistics, computer science, physics, neurobiology, electrical engineering, statistical inference, natural language processing, cryptography, neurobiology, human vision, the evolution and function of molecular codes (bioinformatics), model selection in statistics, thermal physics, quantum computing, linguistics, plagiarism detection, pattern recognition, anomaly detection, gambling, music composition….

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

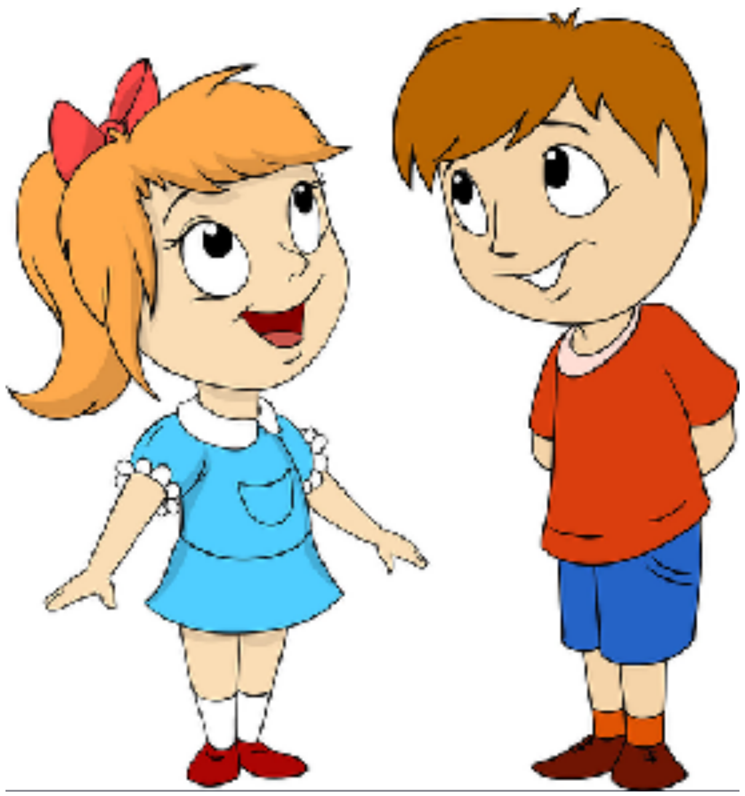
- Information Theory to quantify information

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

$$\begin{array}{cccc} \uparrow & \uparrow & & \uparrow \\ P_1 & P_2 & \ldots & P_n \end{array}$$

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? Pizza or burger?

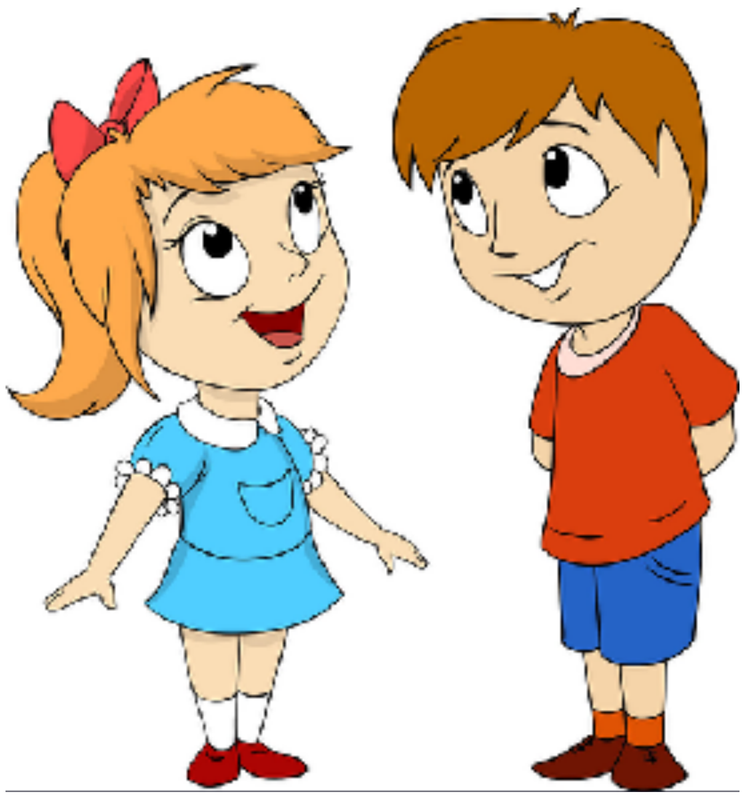
Chantal: Let's have burger!

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

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? Pizza or burger?

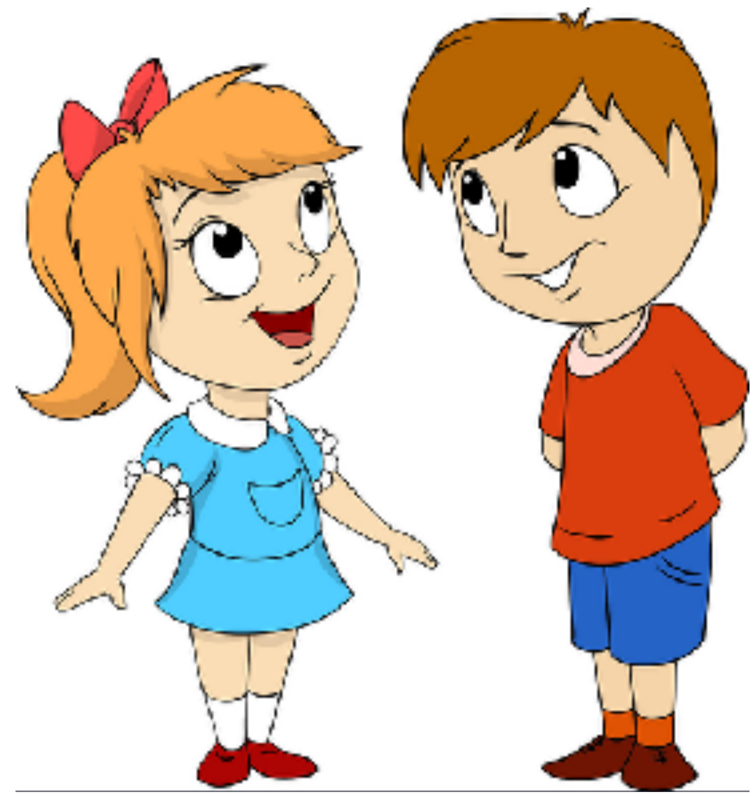
Chantal: Let's have pizza!

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

Before: H(X) = - (0.2 log 0.2 + 0.8 log 0.8) = 0.72 bit

After: H(X)' = 0

- Information Theory to quantify information

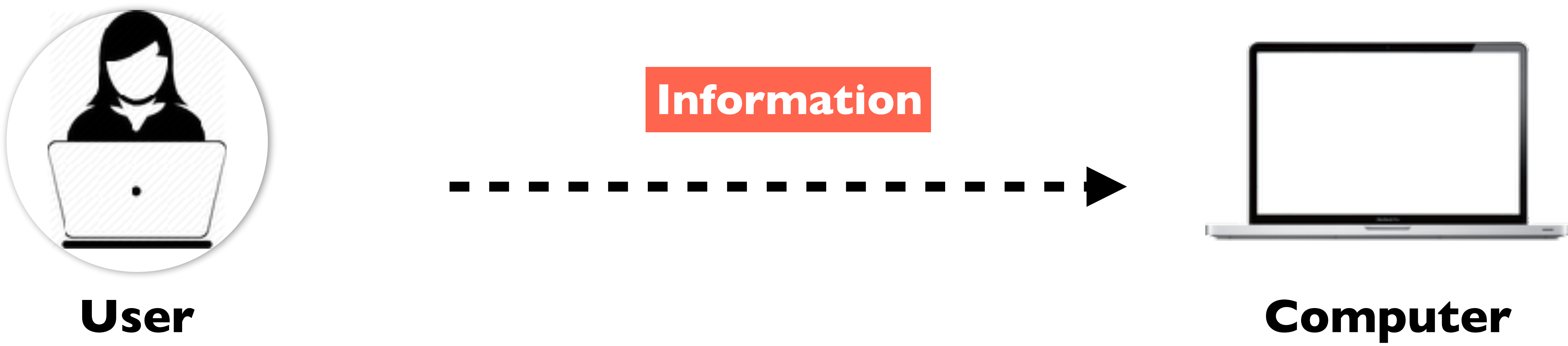


Jean-Pierre: Where do we go for dinner tonight? Pizza or burger?

Chantal: I don't know. We had pizza the other day so I would feel more like burger tonight. But I really like pizza. 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

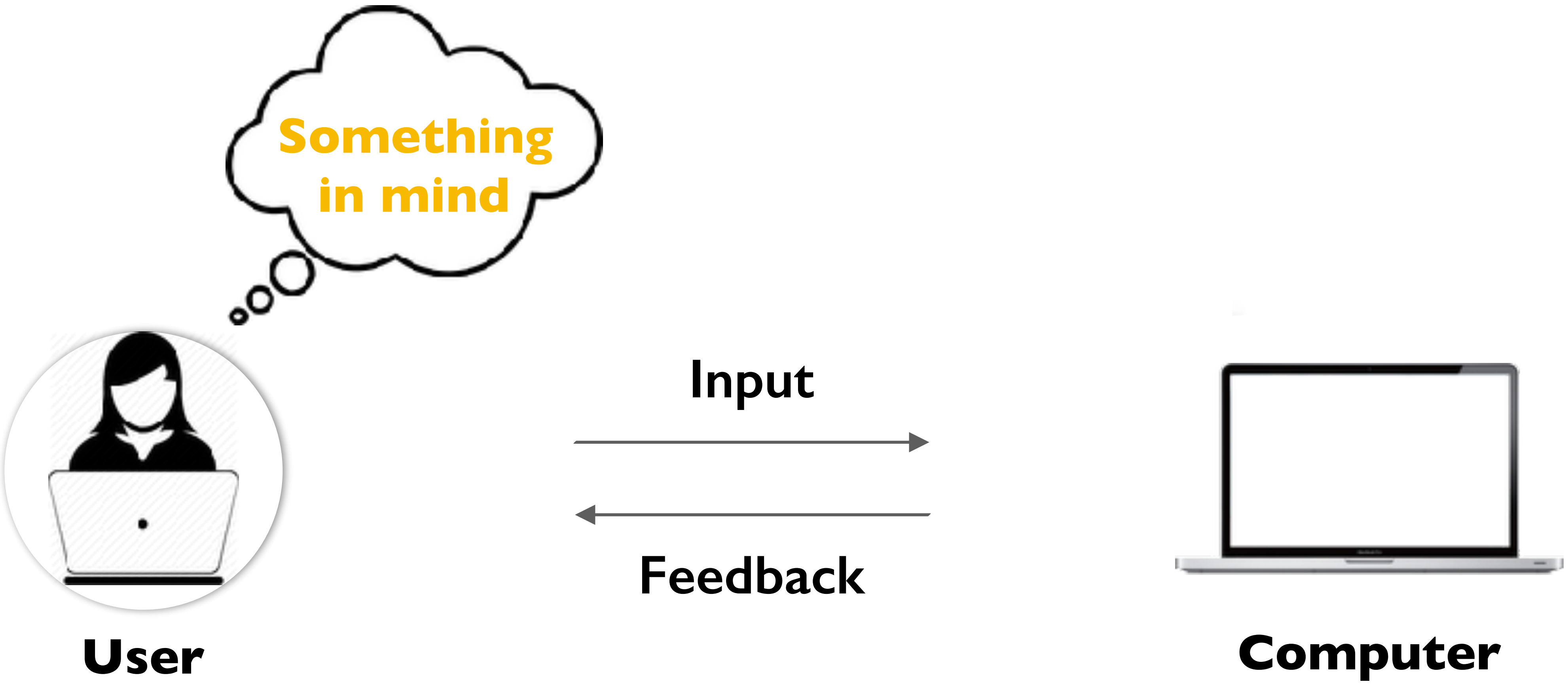
Commands

Any input?

- Bayesian Information Gain (BIG)

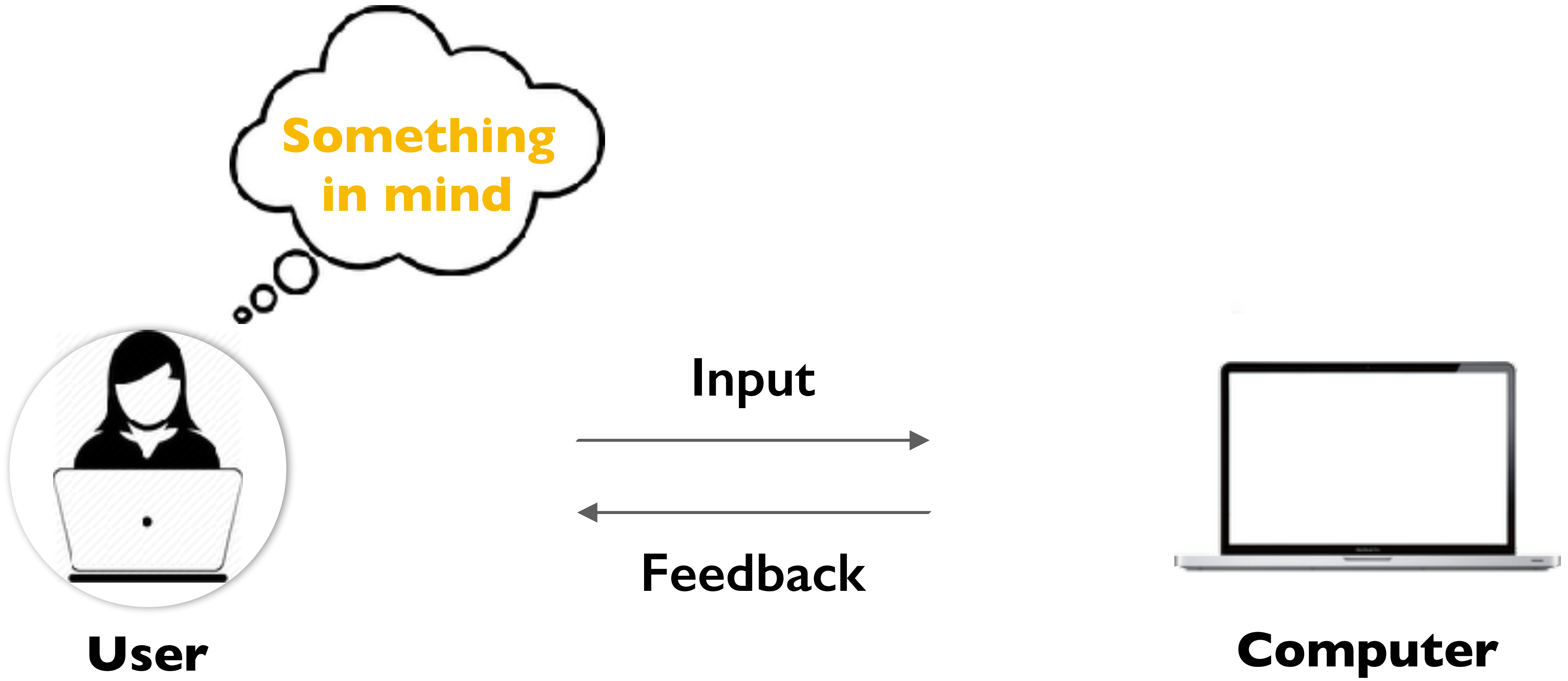
The computer gains **information** from the **user input** to reduce its **uncertainty** about the user’s goal.

- Bayesian Information Gain (BIG)



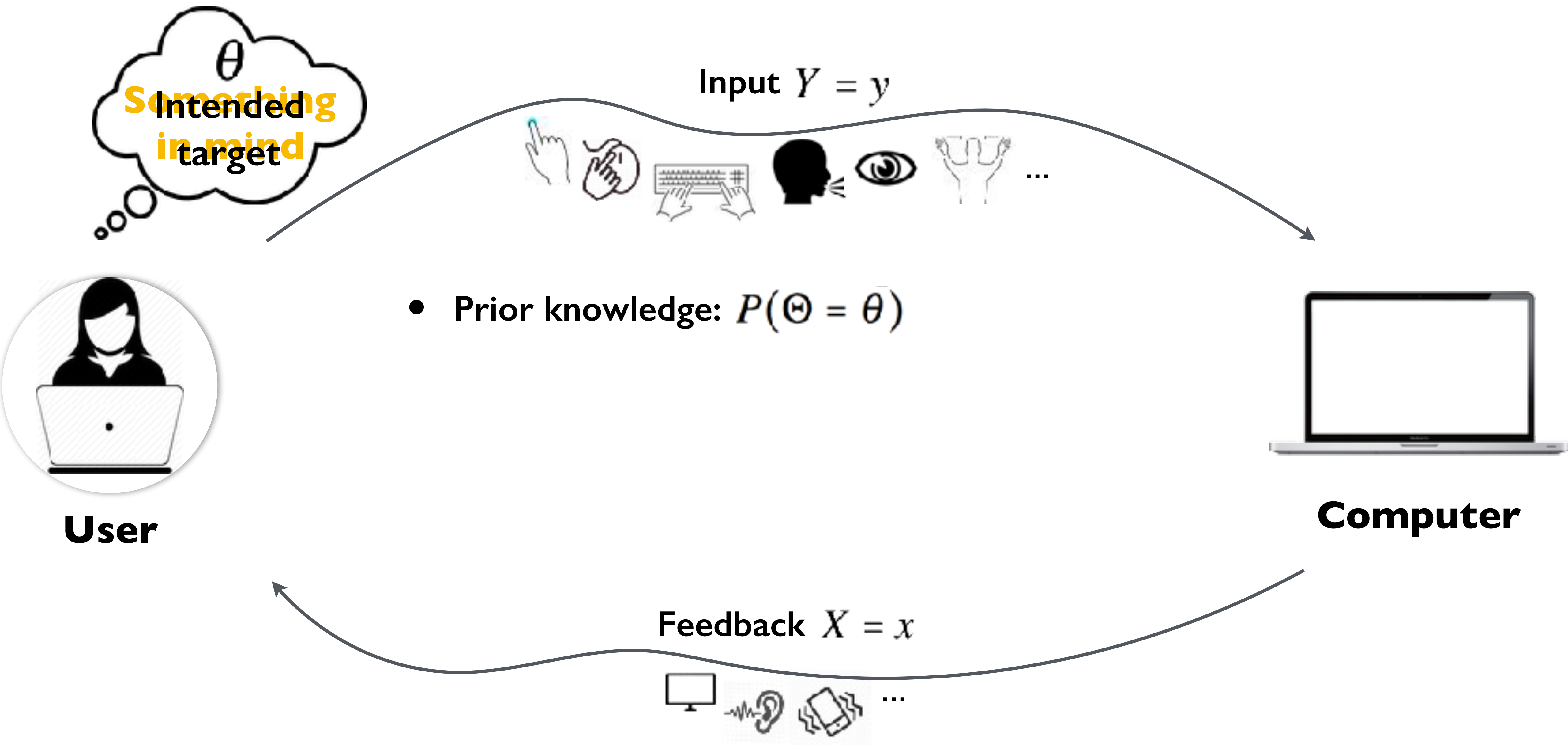
- To look for a restaurant
- To type a word
- To draw a gesture
- To select an icon
- To do something
- ....

- Bayesian Information Gain (BIG)



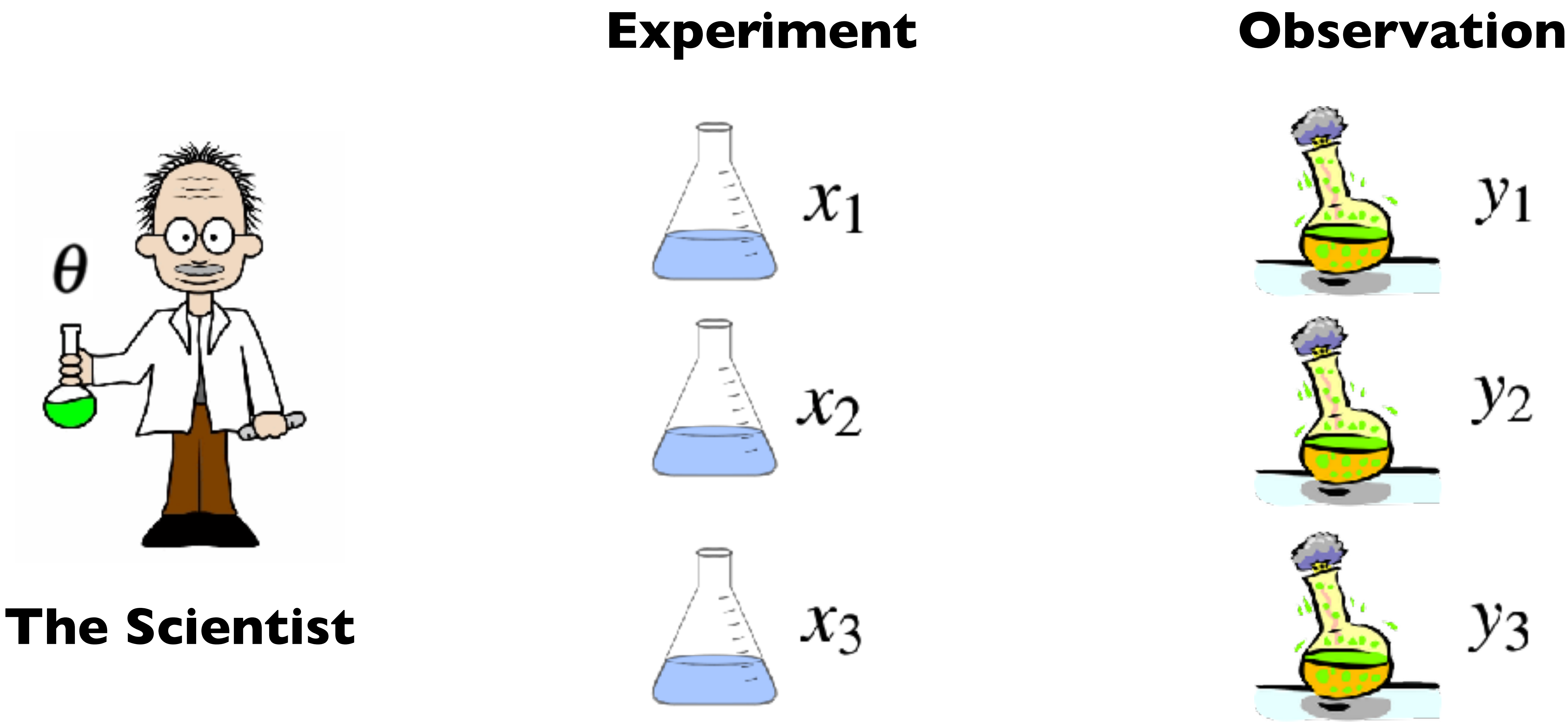
- Uncertainty about this something
- Uncertainty reduces gradually when receiving user input

- Bayesian Information Gain (BIG)



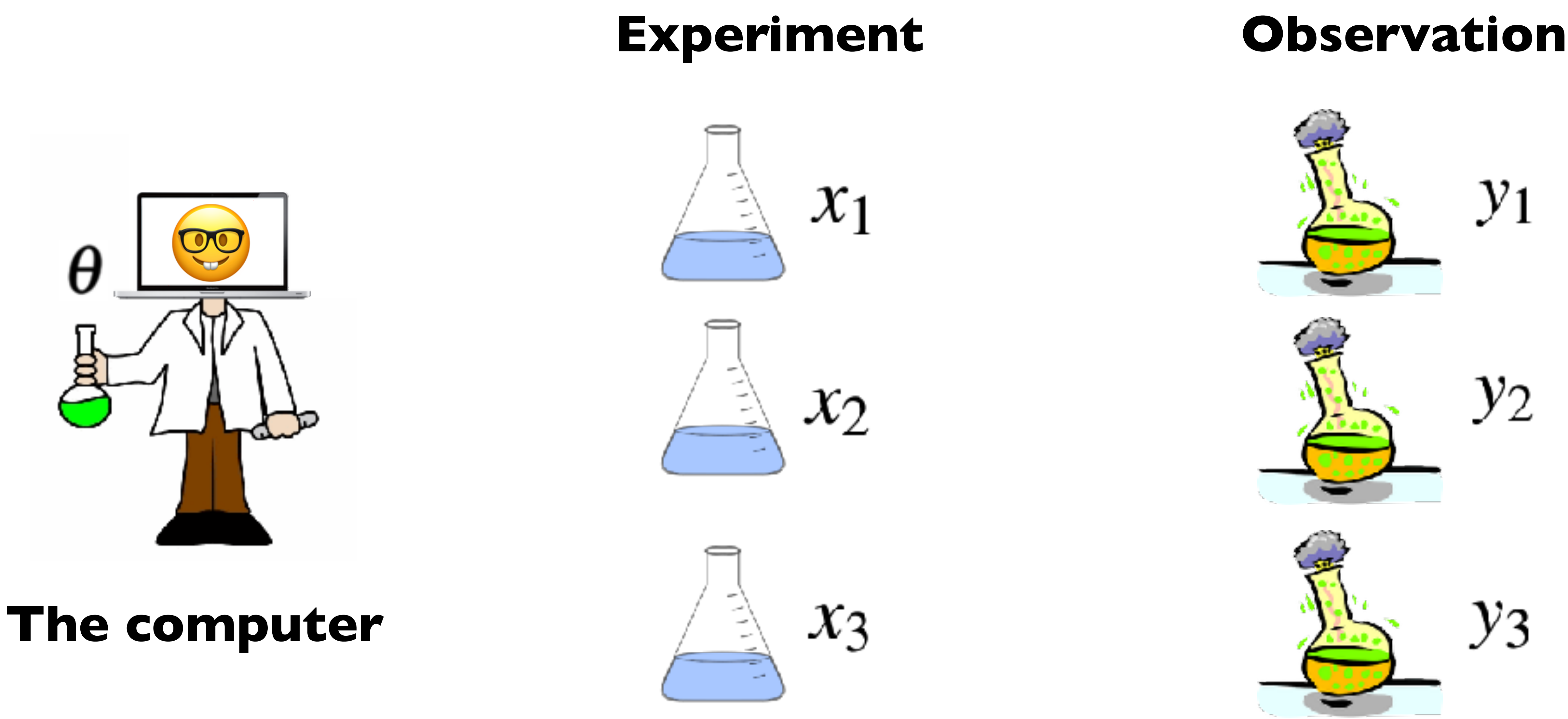
- Prior knowledge: $P(\Theta = \theta)$

- Bayesian Information Gain (BIG)



On a Measure of the Information Provided by an Experiment (Lindley 1956)

- Bayesian Information Gain (BIG)



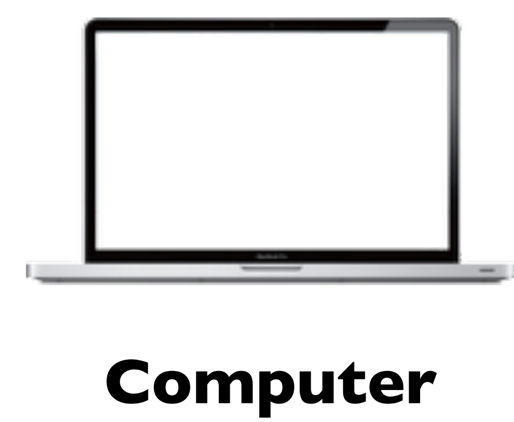
On a Measure of the Information Provided by an Experiment (Lindley 1956)

- 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, Entrain**

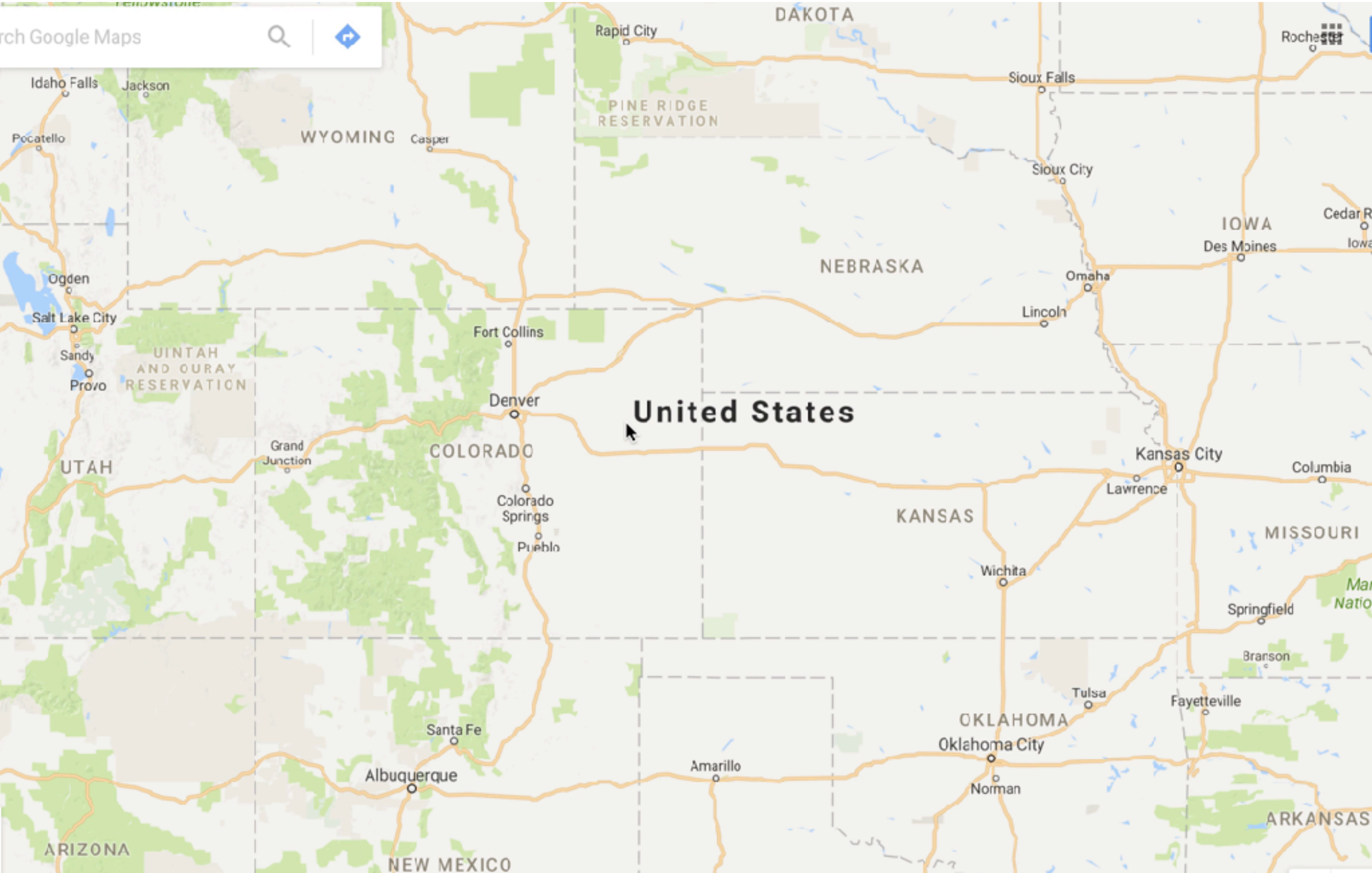
Input $Y = y$

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

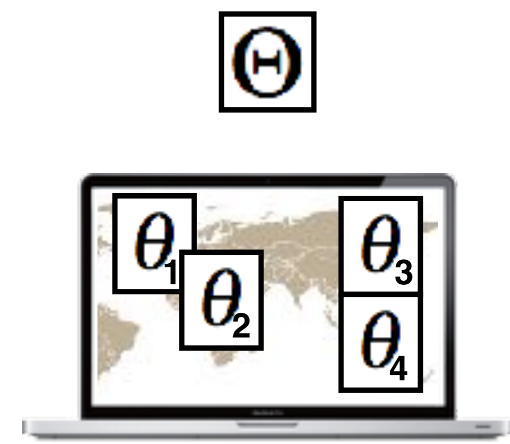


Feedback $X = x$

- Bayesian Information Gain (BIG) - Standard navigation

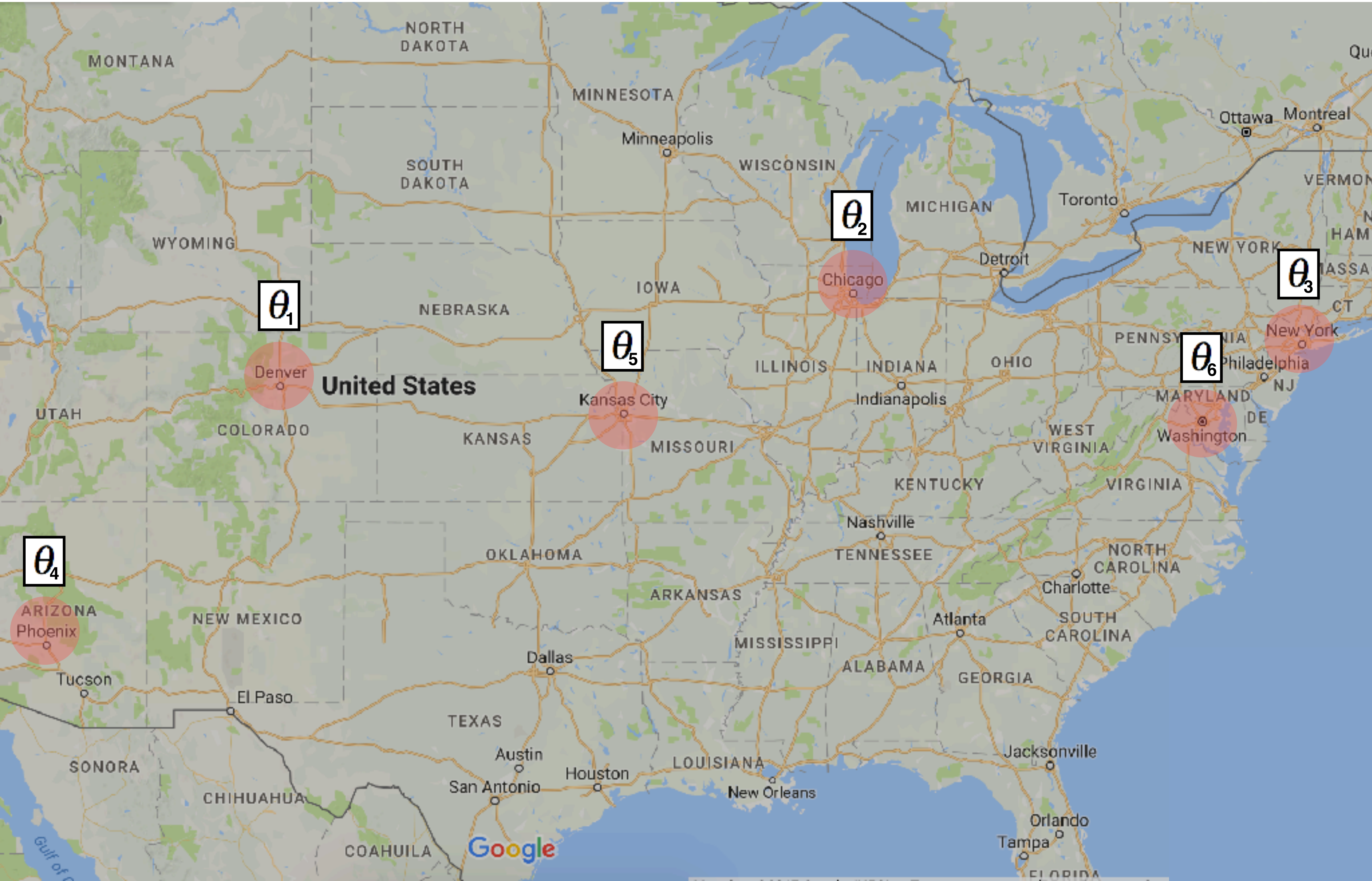


- Bayesian Information Gain (BIG) - Standard navigation



$$P(\Theta = \theta_i)$$

- Bayesian Information Gain (BIG) - Standard navigation

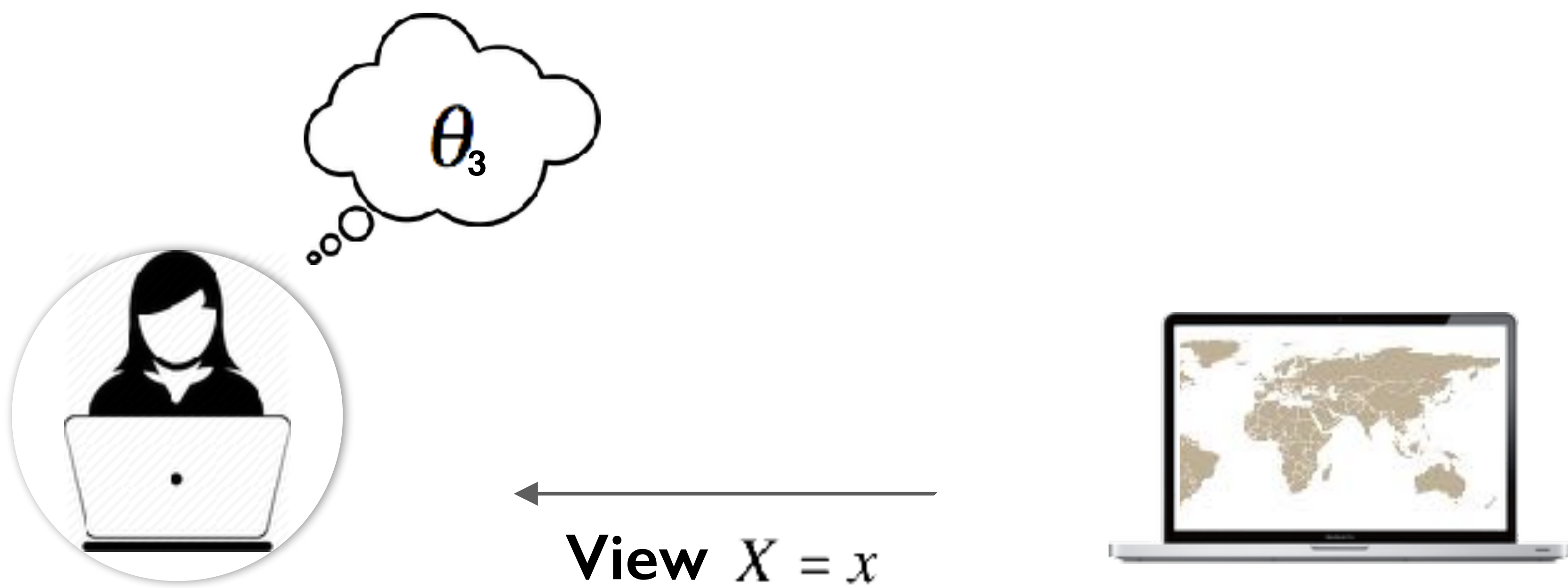


- Bayesian Information Gain (BIG) - Standard navigation

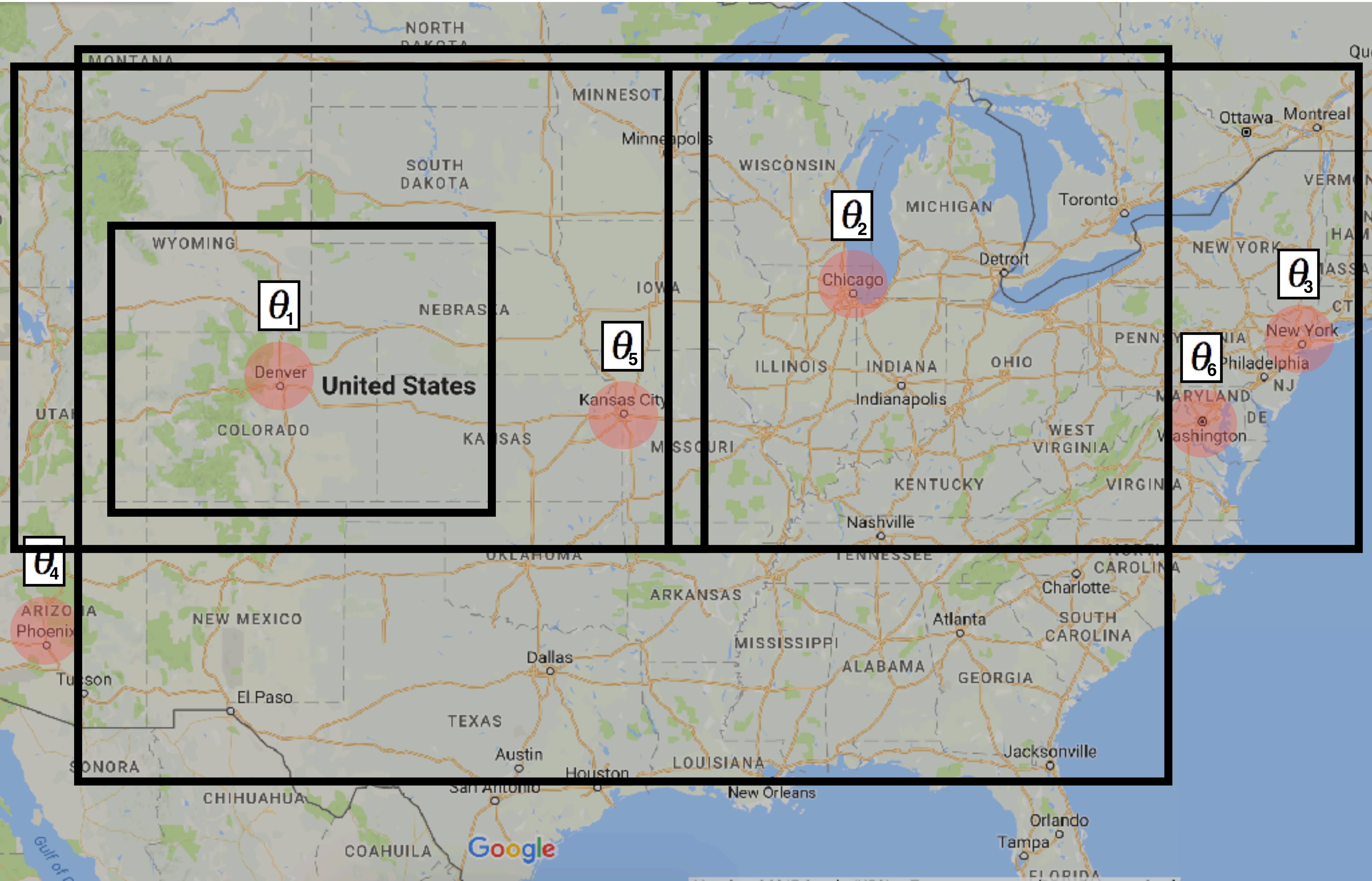
* **The computer's Uncertainty about the user's goal**

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

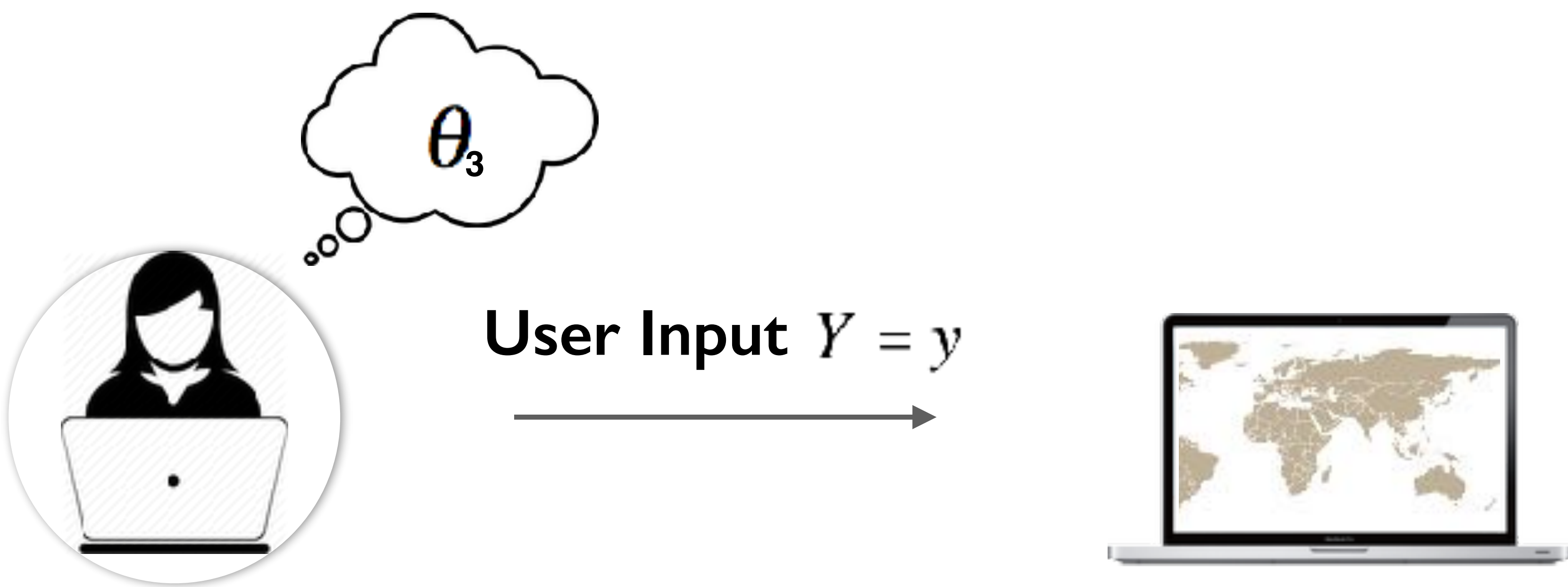
- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation

- Bayesian Information Gain (BIG) - Standard navigation

$$P(Y = y | \Theta = \theta, X = x)$$

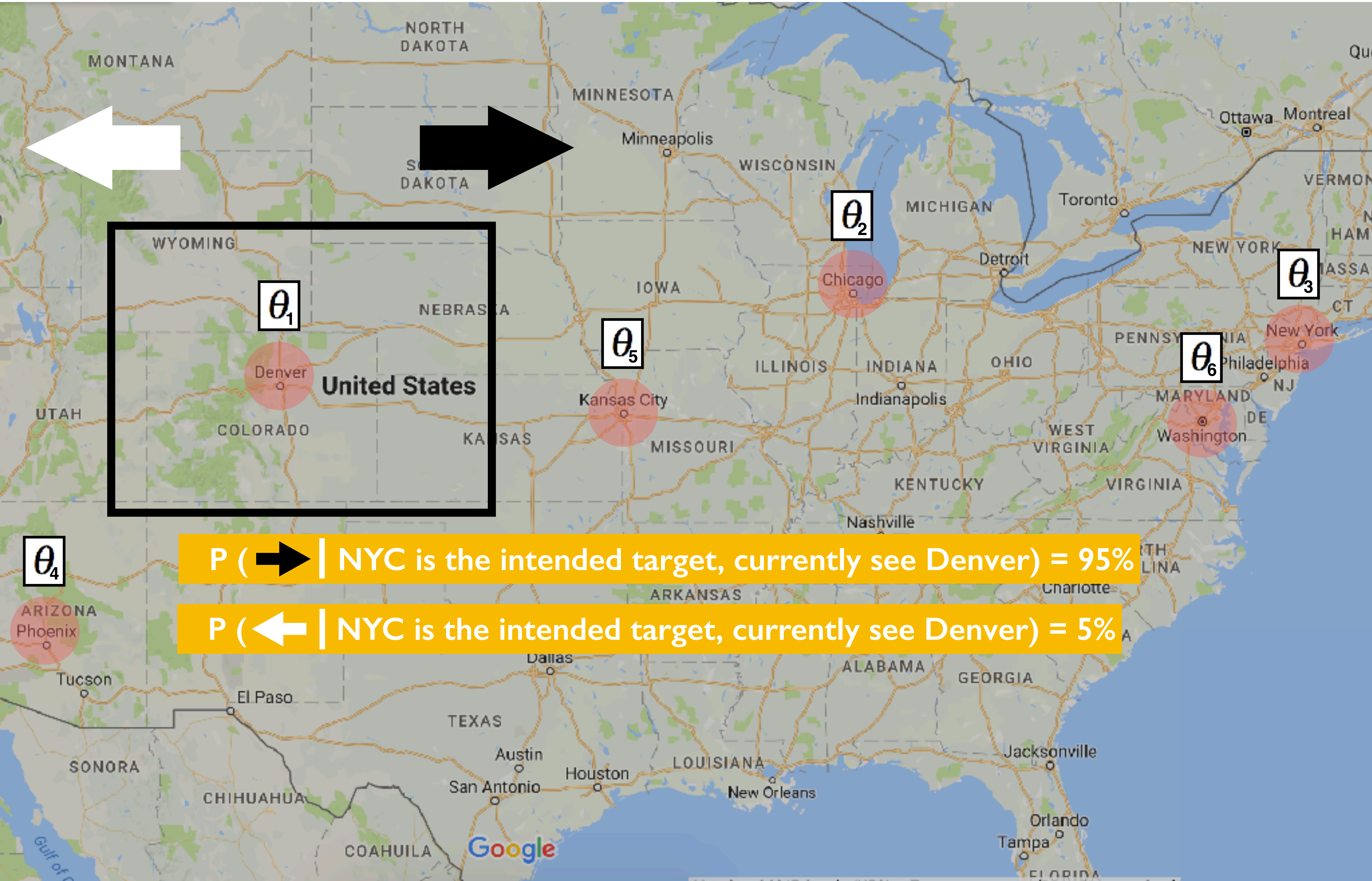
Interpret User Input

$$P(\Theta = \theta_i)$$

View $X = x$

User Input $Y = y$

- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation

View $X = x$

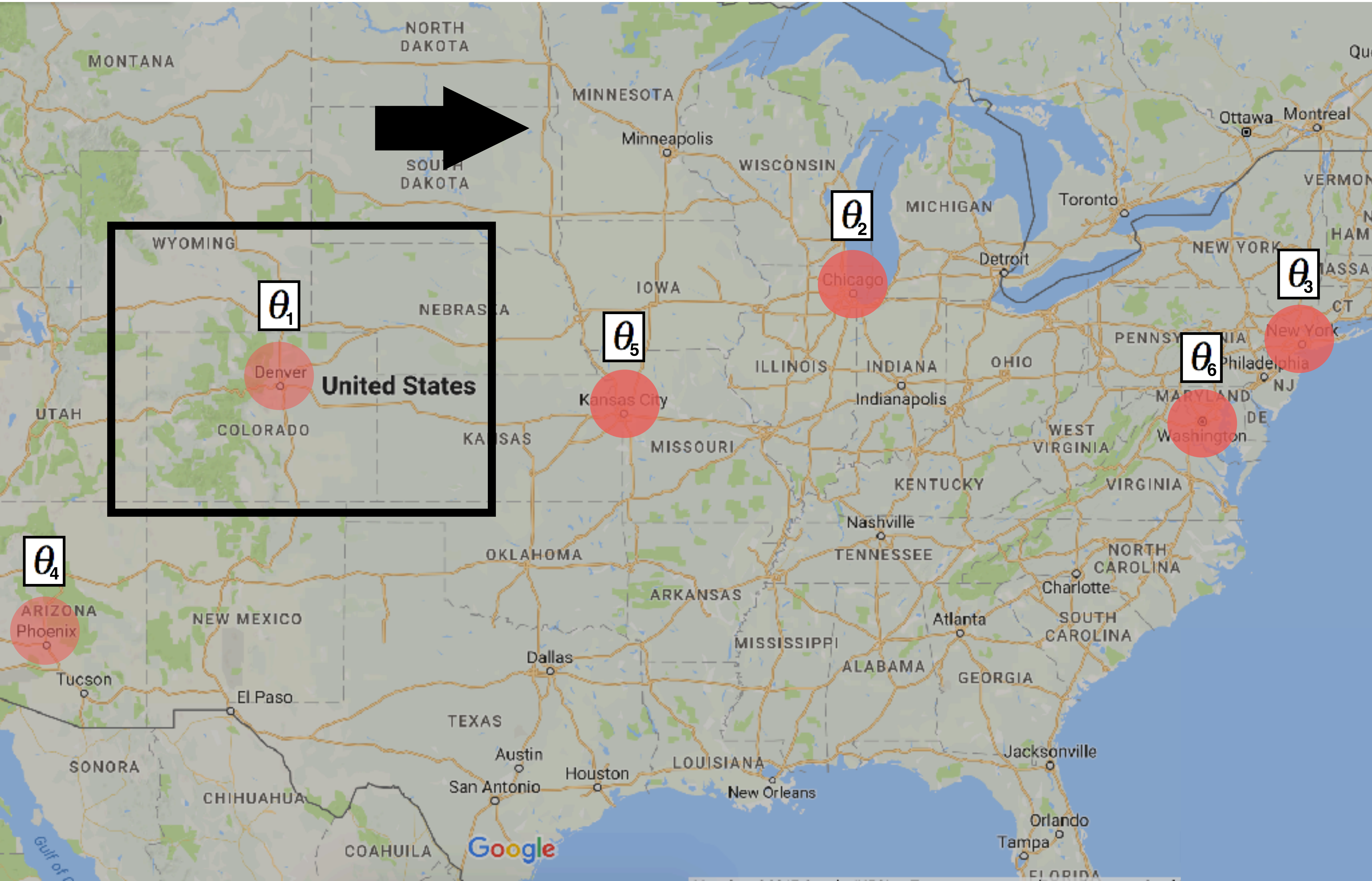
$$P(\Theta = \theta \mid X = x, Y = y)$$

Update its knowledge

$$P(\Theta = \theta_i)$$

User Input $Y = y$

- Bayesian Information Gain (BIG) - Standard navigation



- Bayesian Information Gain (BIG) - Standard navigation

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

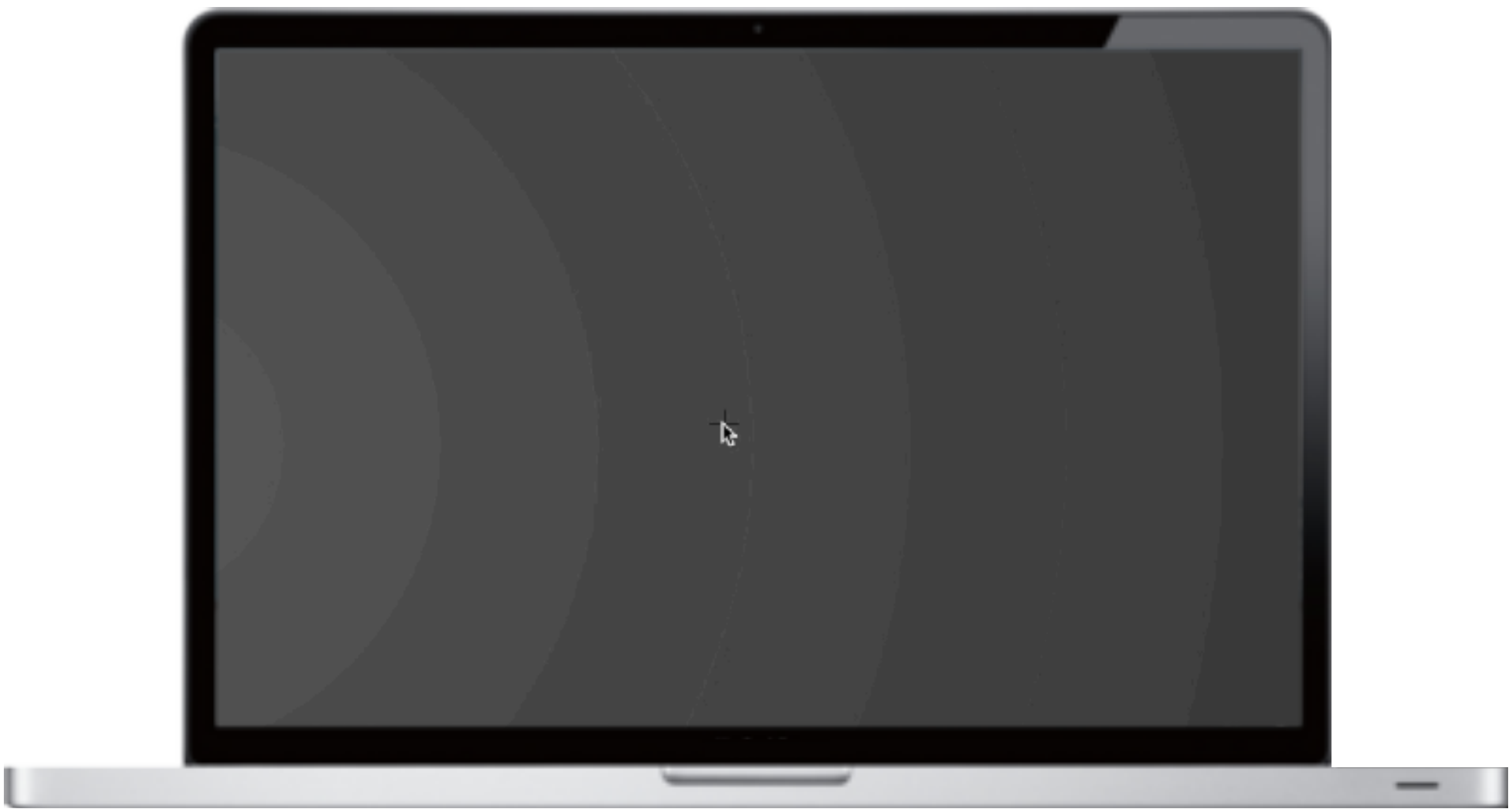
$$H(\Theta) = -\sum_{i=1}^{n} P(\Theta = \theta_i) \log_2 P(\Theta = \theta_i)$$

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

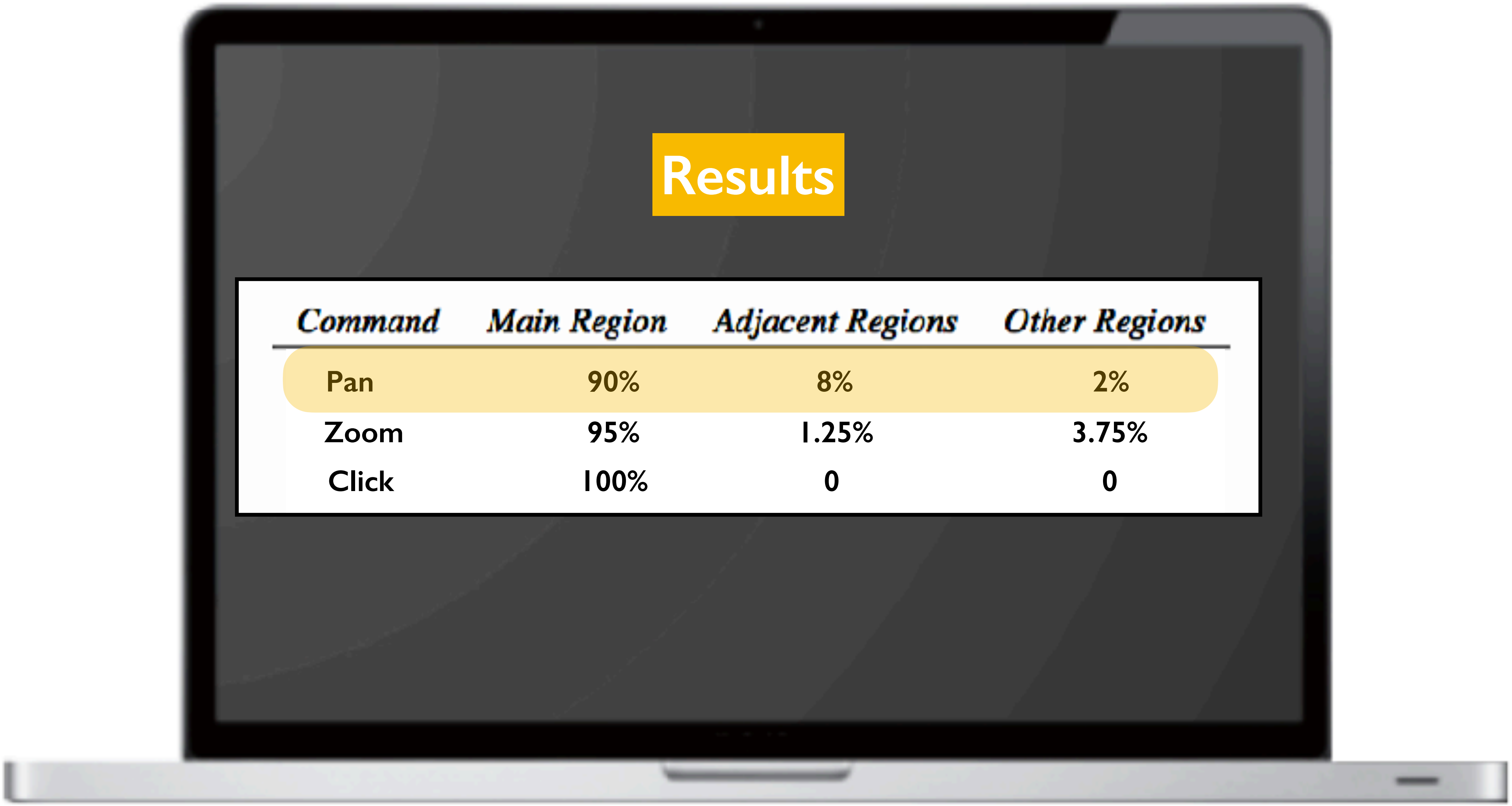
$$P(\Theta = \theta \mid X = x, Y = y) = \frac{P(Y = y \mid \Theta = \theta, X = x) P(\Theta = \theta)}{P(Y = y \mid X = x)}$$

- 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



Results

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

* **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**

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

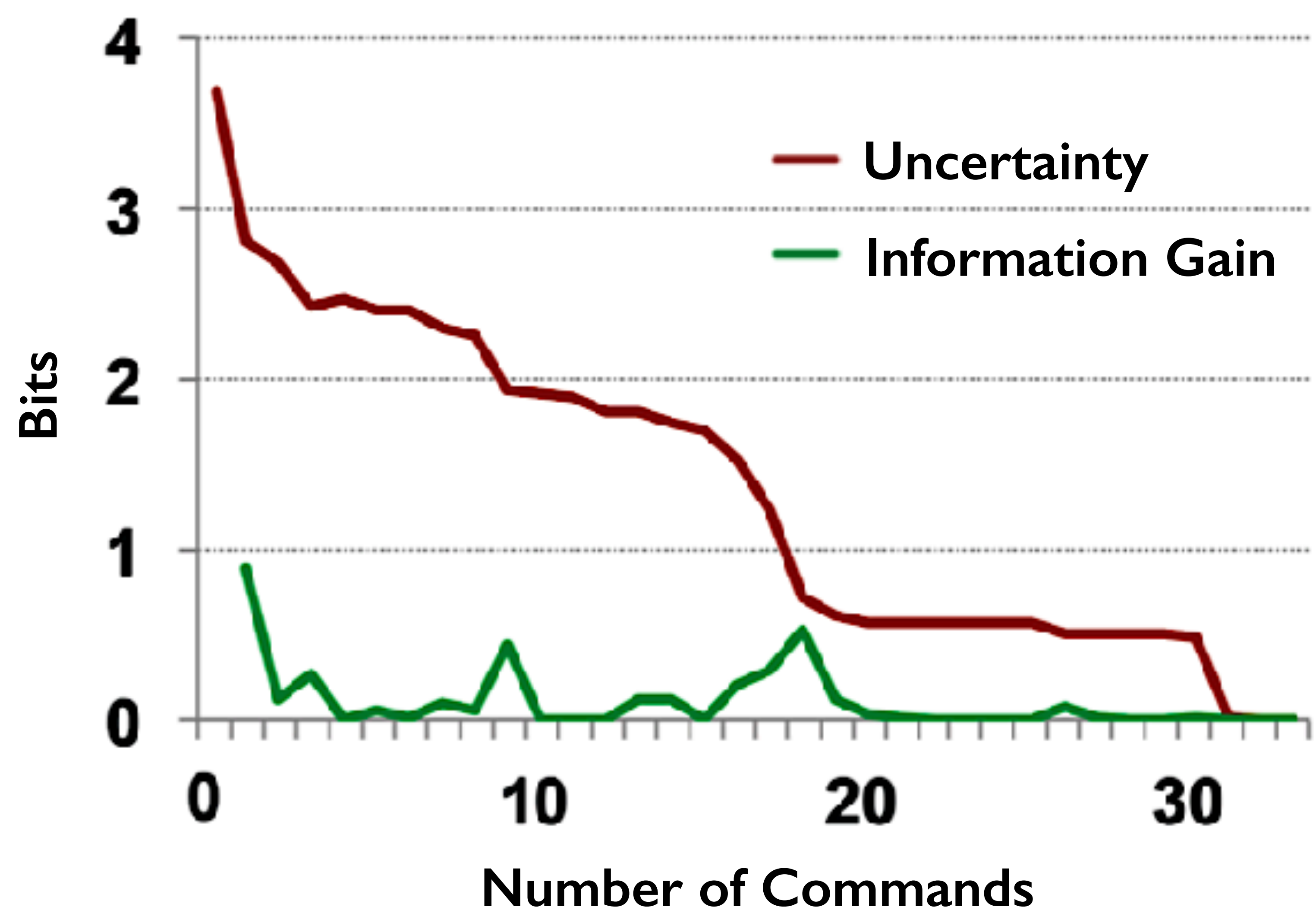
* **The information in the user's input for reducing the computer's uncertainty**

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

- Bayesian Information Gain (BIG) - Standard navigation

✻ Executes the user input only Multiscale navigation

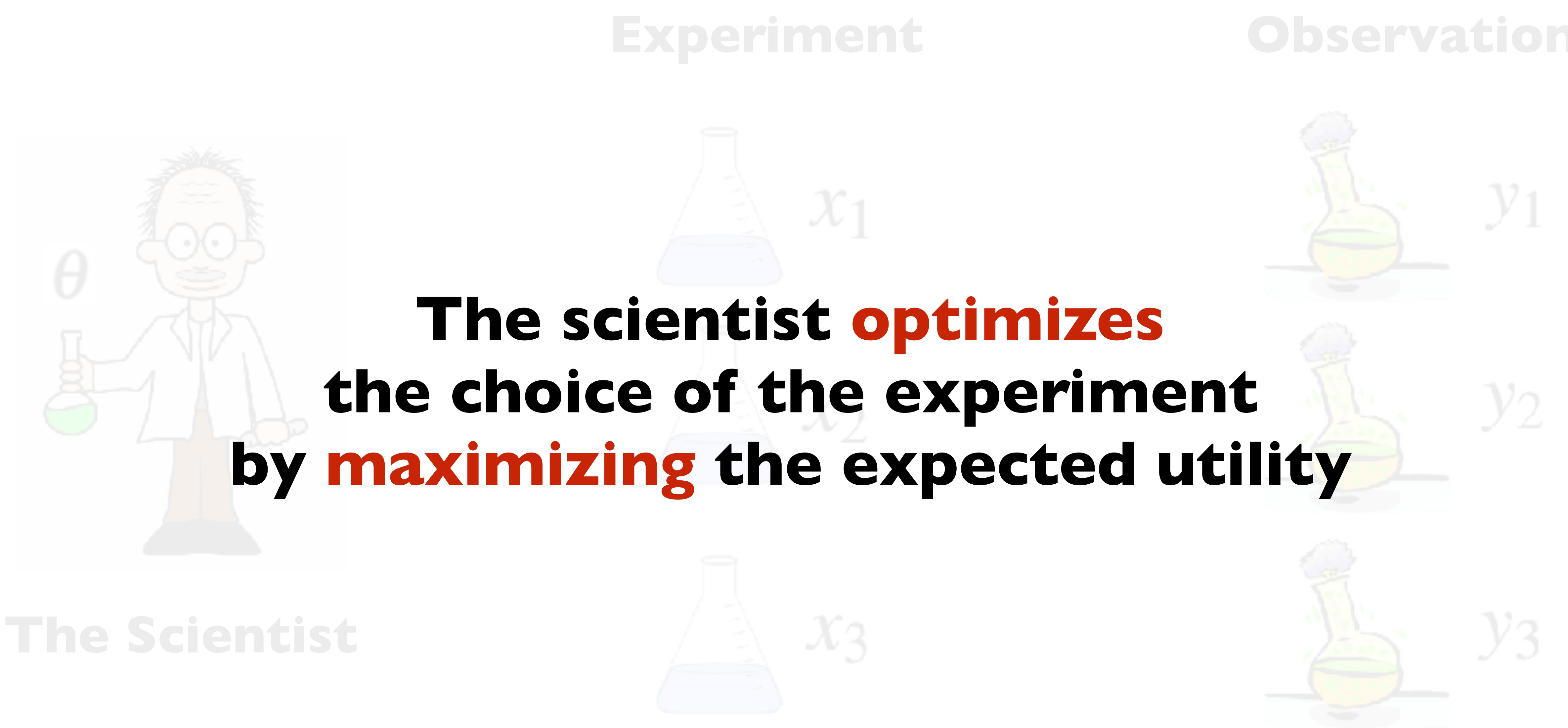
✻ Each user input does not provide much information for the computer to know her goal



- Bayesian Information Gain (BIG) - BIGnav

Can we challenge users to give more information?

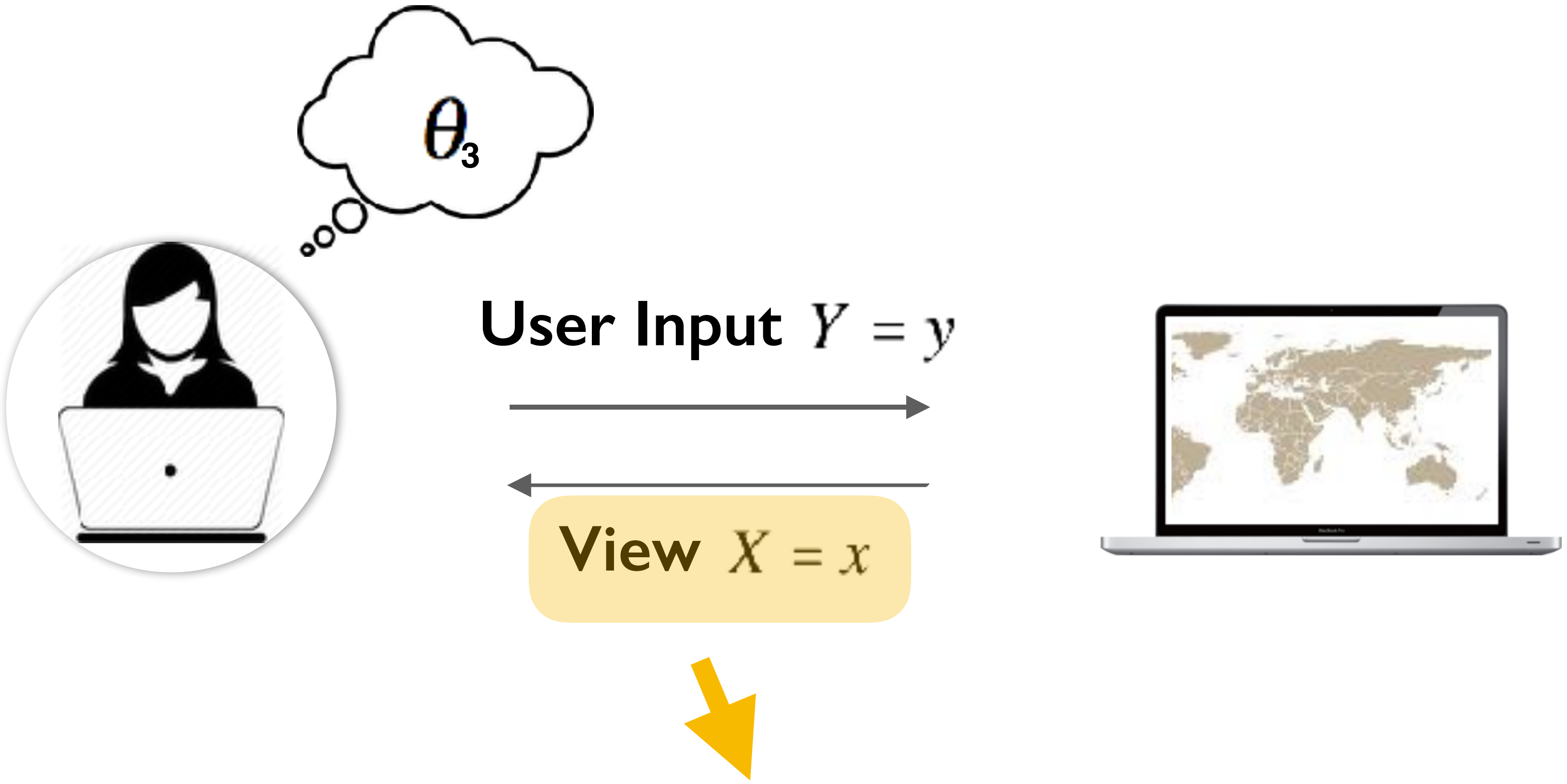
- Bayesian Information Gain (BIG) - BIGnav



The scientist optimizes
the choice of the experiment
by maximizing the expected utility

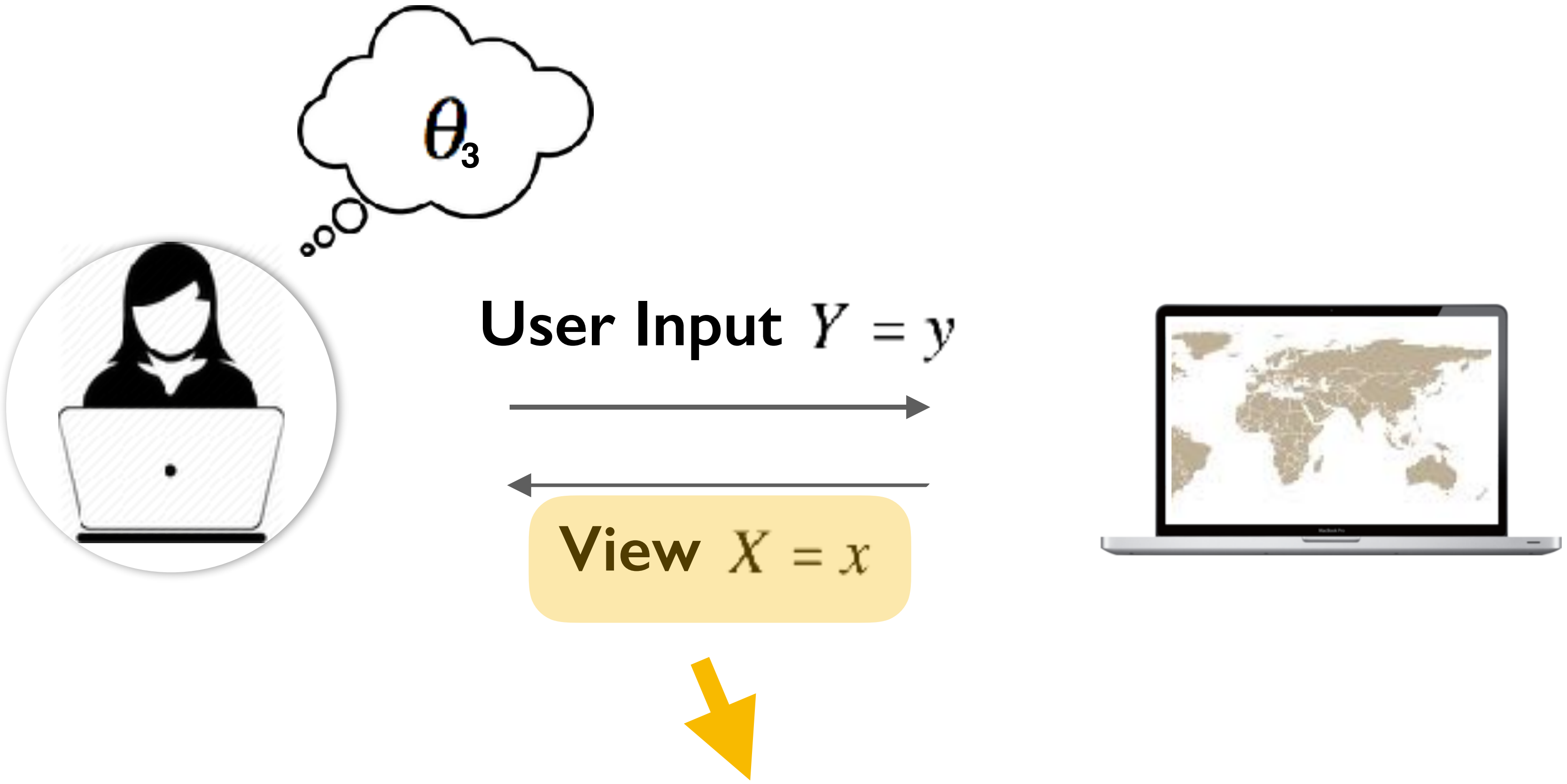
On a Measure of the Information Provided by an Experiment (Lindley 1956)

- Bayesian Information Gain (BIG) - BIGnav



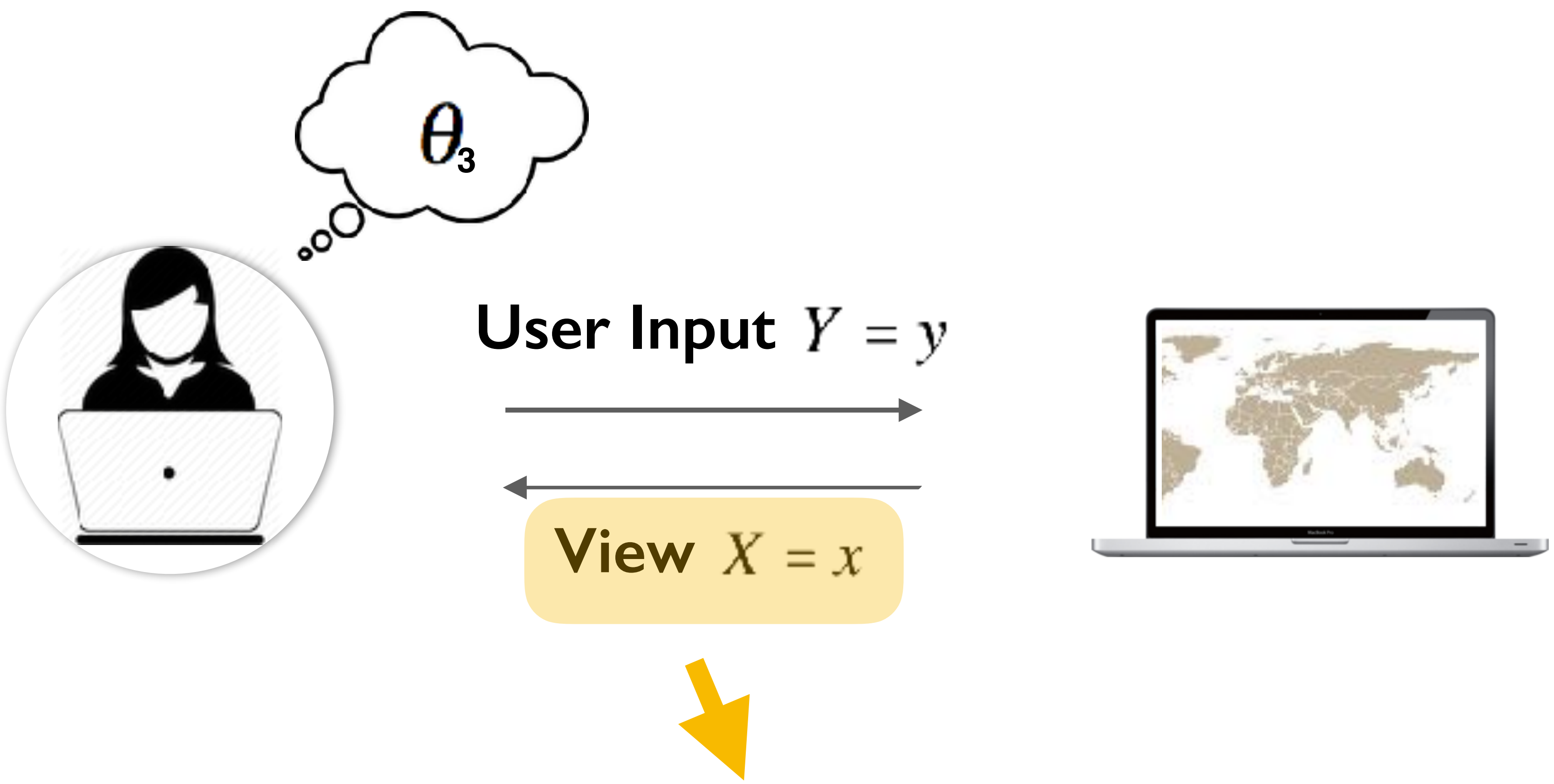
The expected information gain

- Bayesian Information Gain (BIG) - BIGnav



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

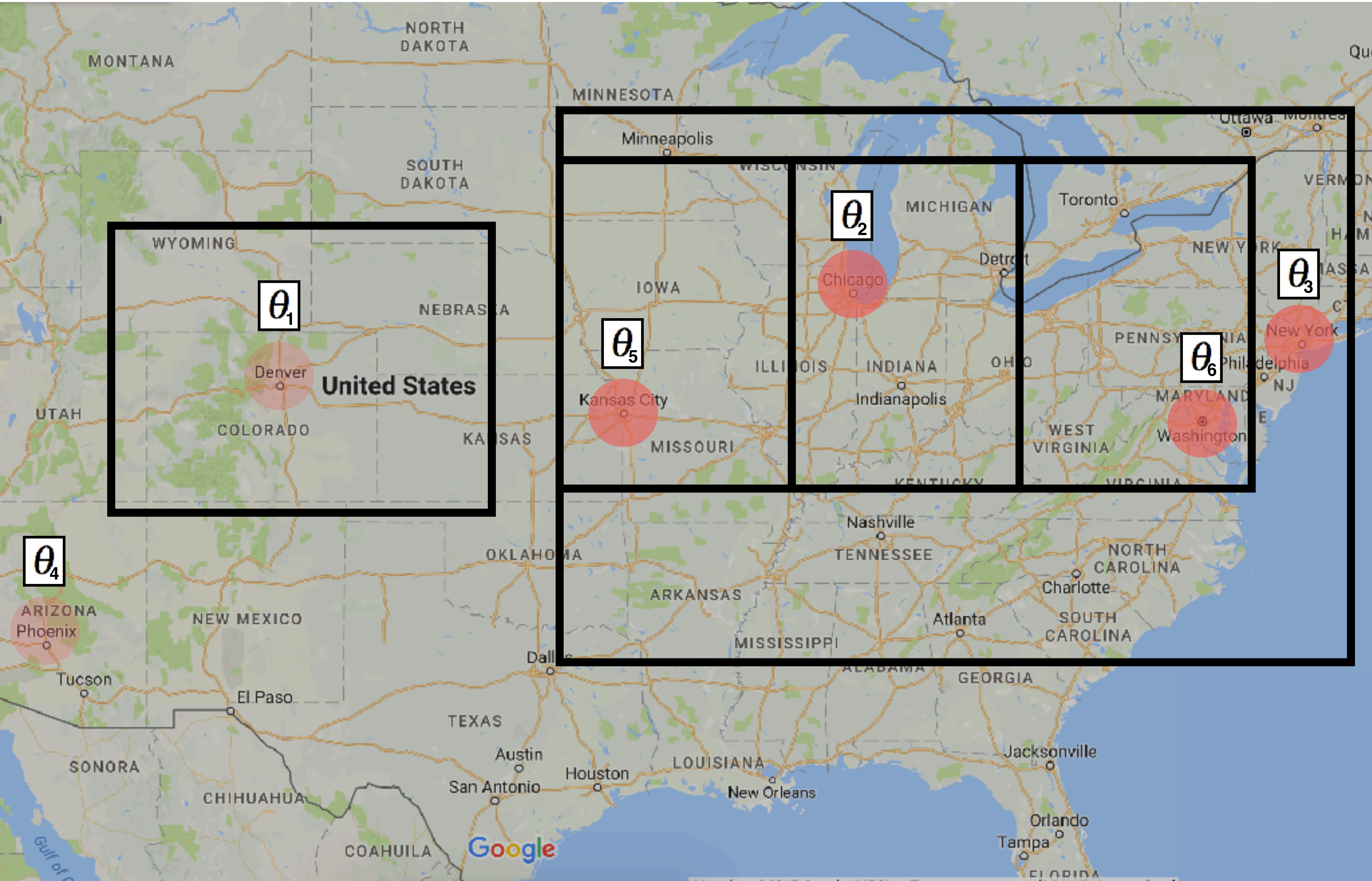
- Bayesian Information Gain (BIG) - BIGnav



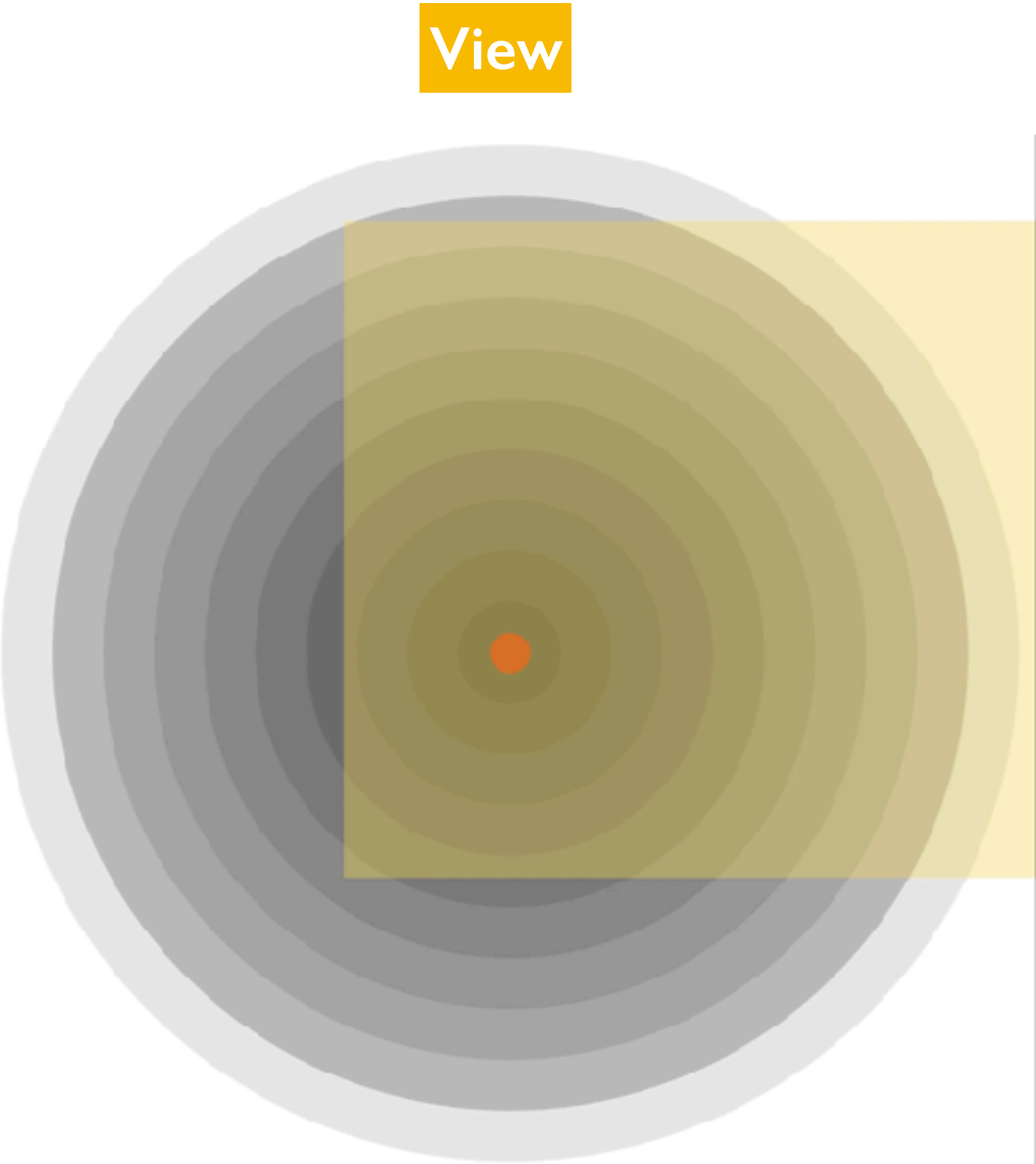
Go over all possible feedback, and find the one that maximizes the expected information gain

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

- Bayesian Information Gain (BIG) - BIGnav

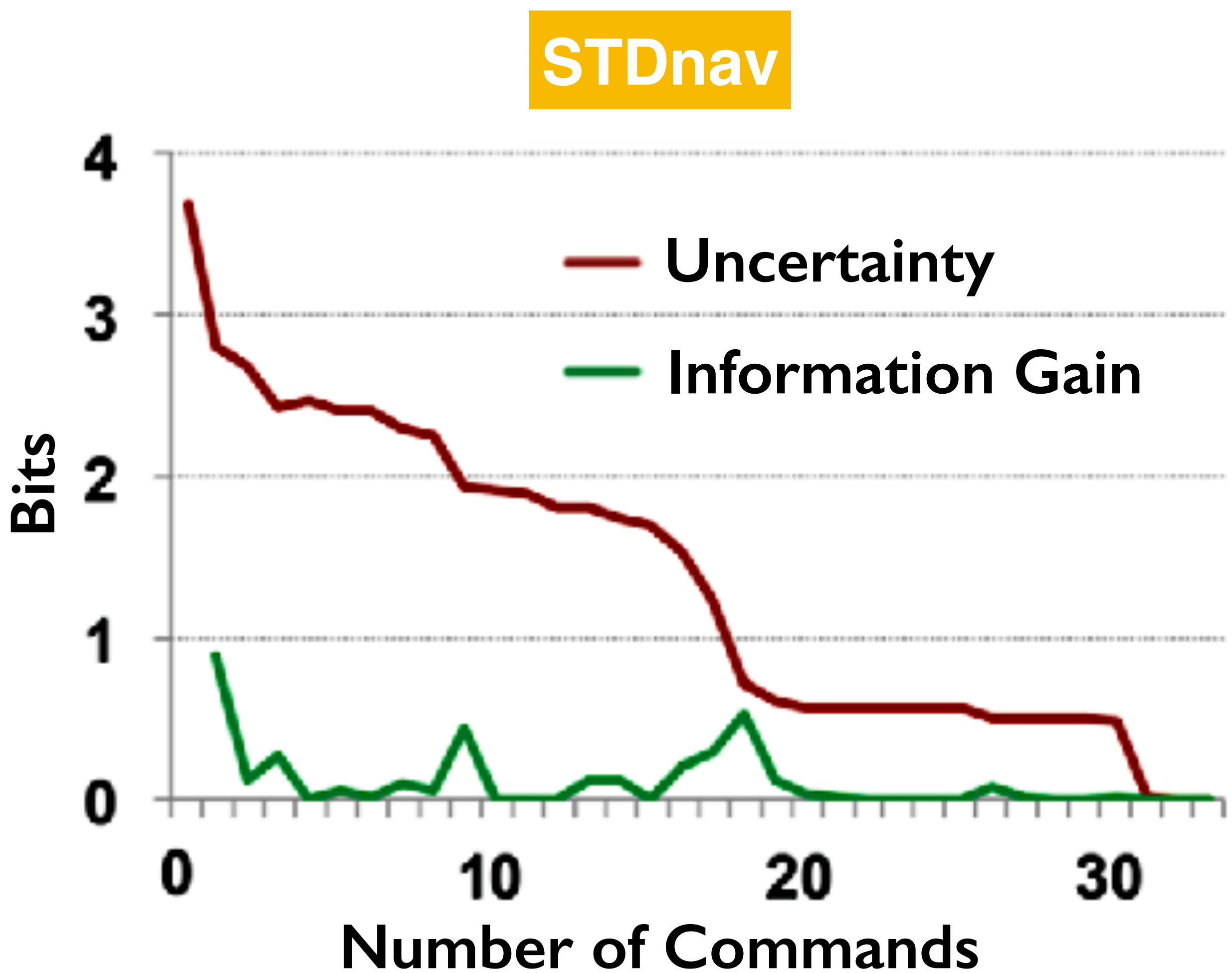


- Bayesian Information Gain (BIG) - BIGnav



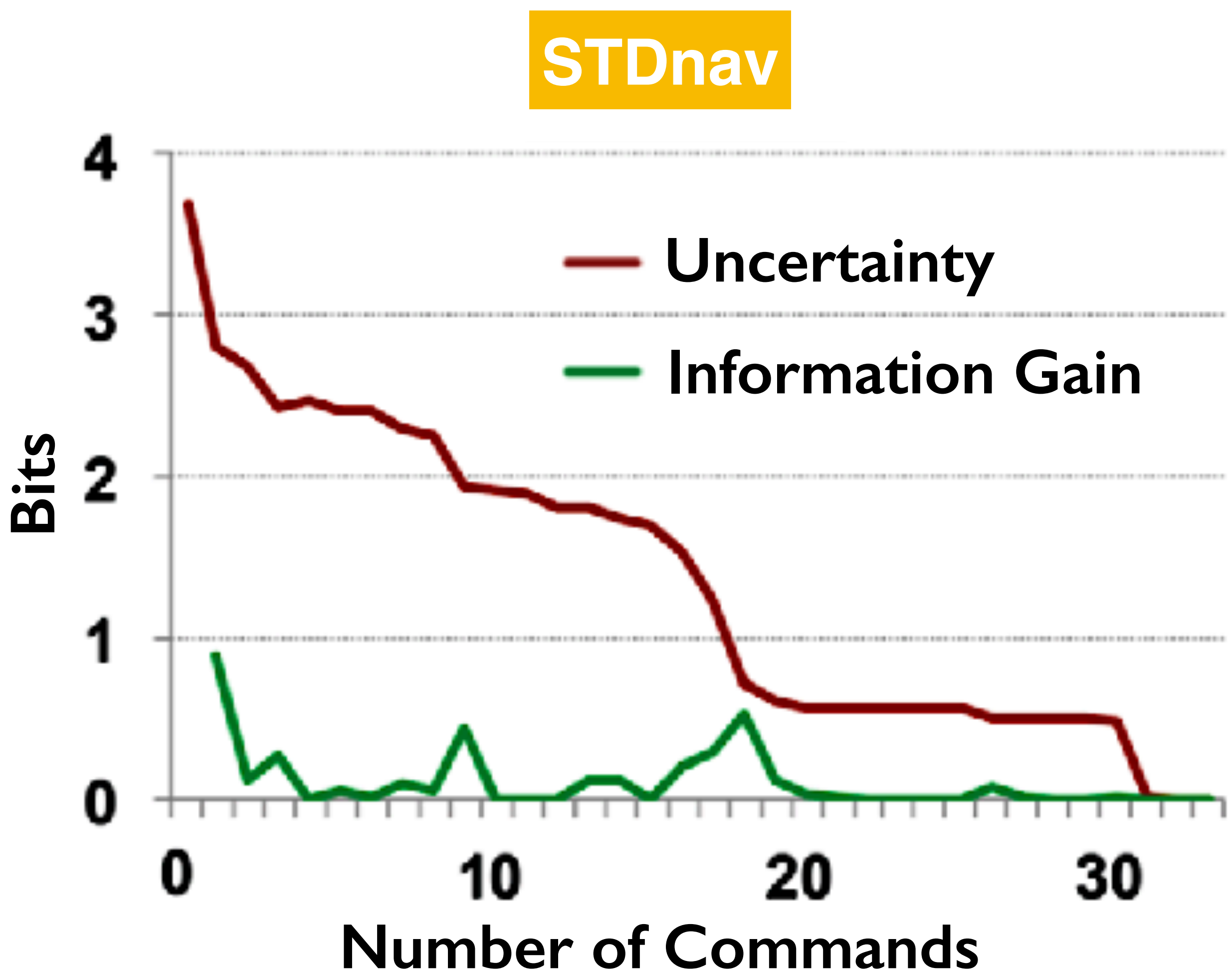
- Bayesian Information Gain (BIG) - BIGnav

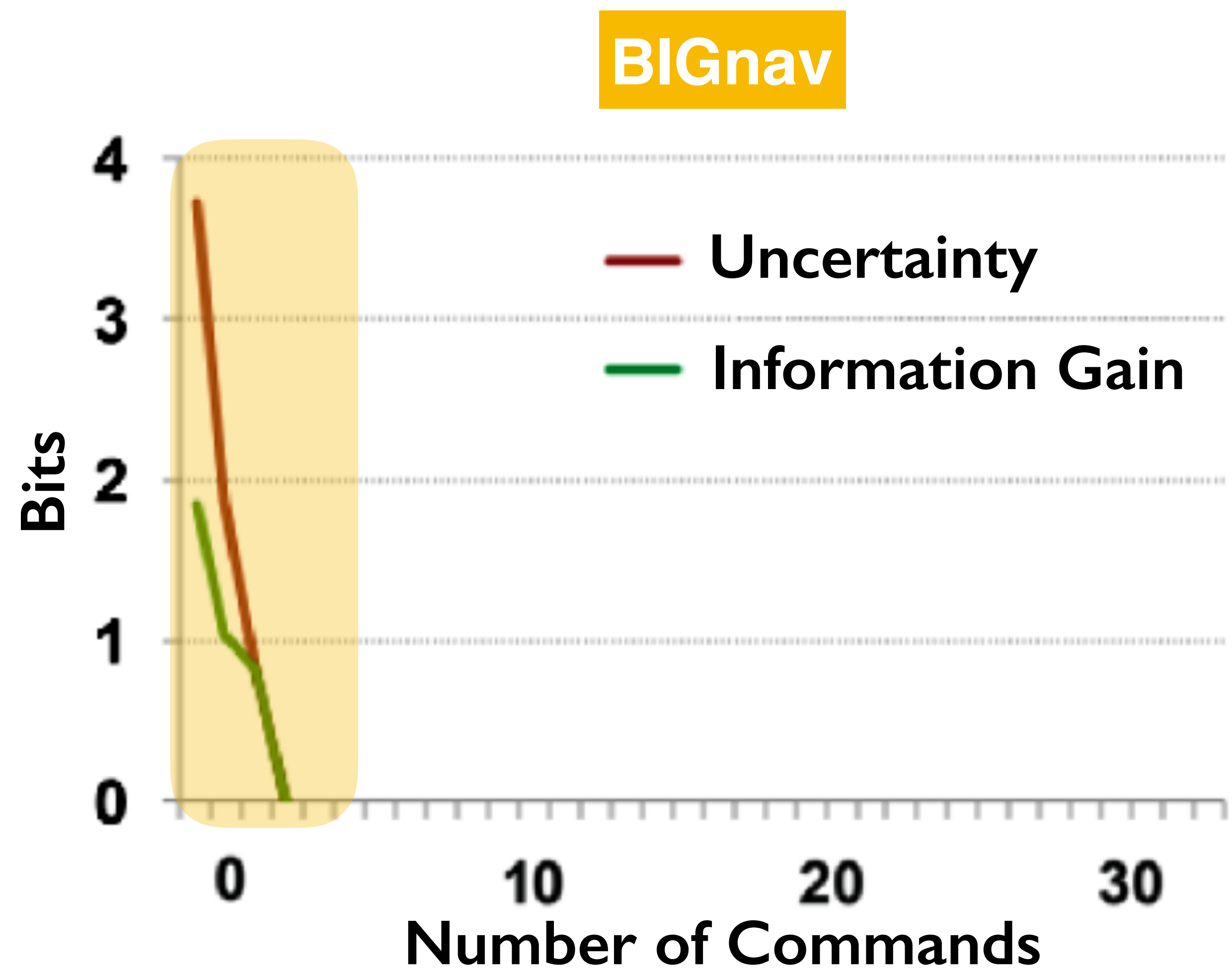
✻ **BIGnav gains maximum information from each user input**



- Bayesian Information Gain (BIG) - BIGnav

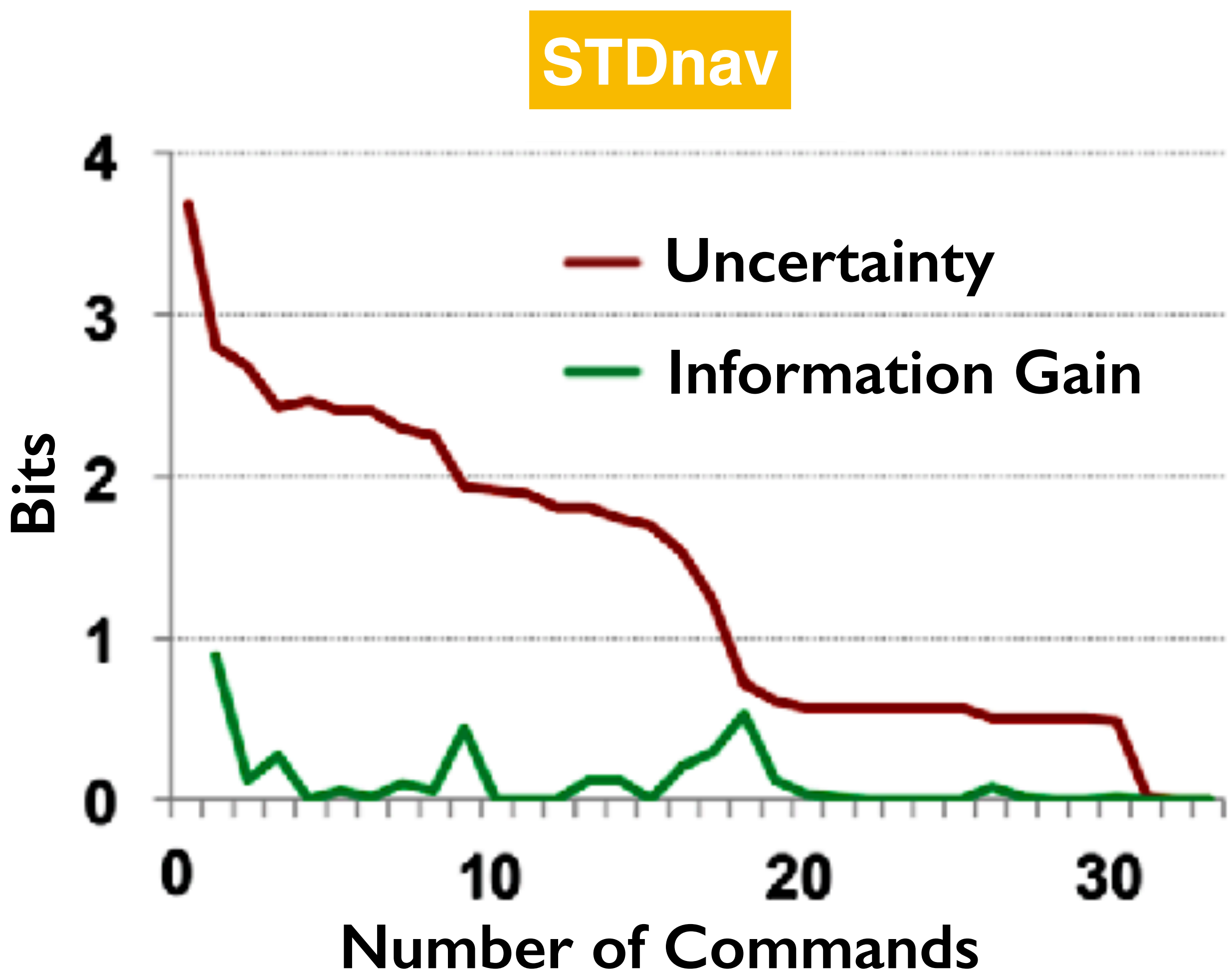
✻ BIGnav gains maximum information from each user input

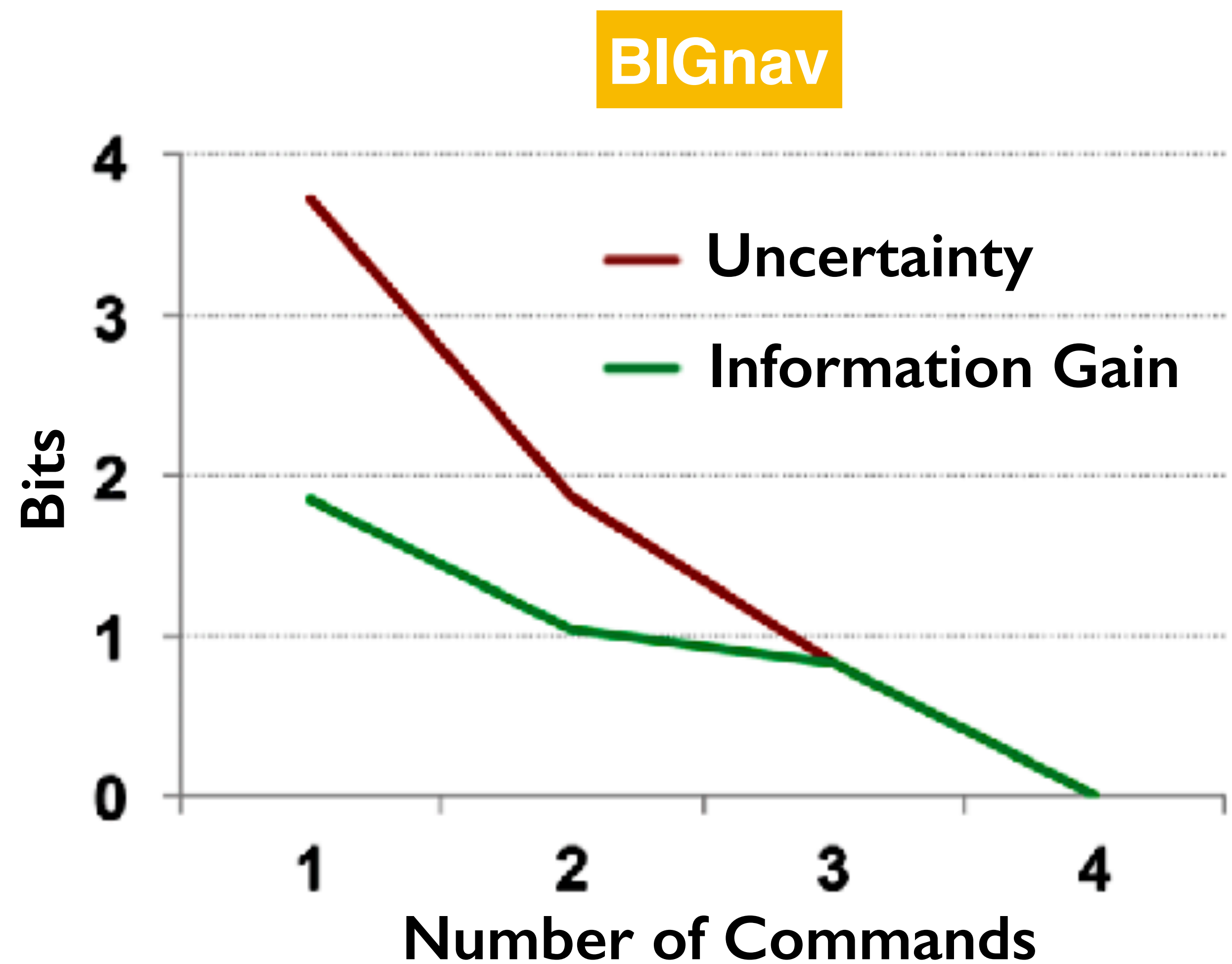




- Bayesian Information Gain (BIG) - BIGnav

✻ 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

Coffee Break

- Bayesian Information Gain (BIG) - BIGnav

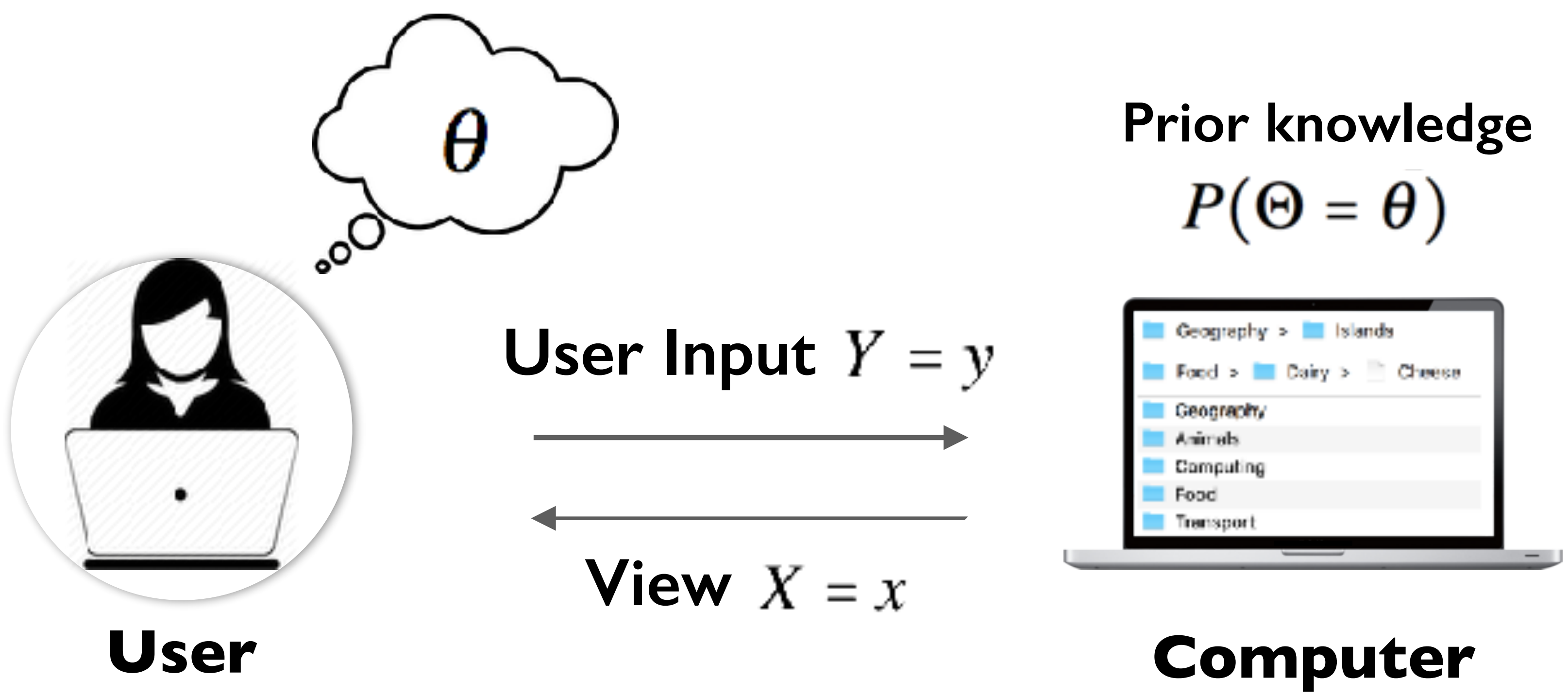
More efficient but…

Higher cognitive load

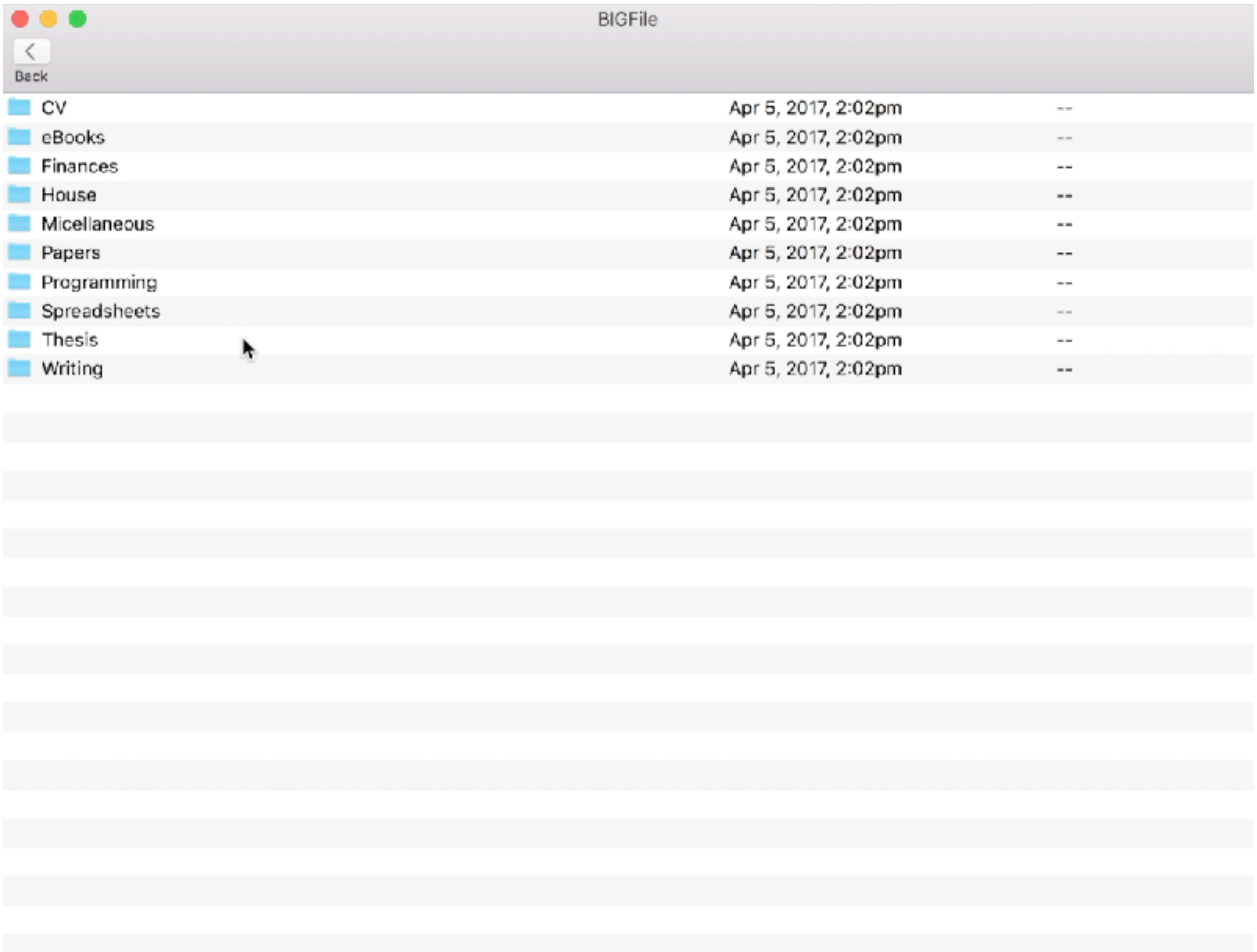
🙄

- Bayesian Information Gain (BIG) - BIGFile

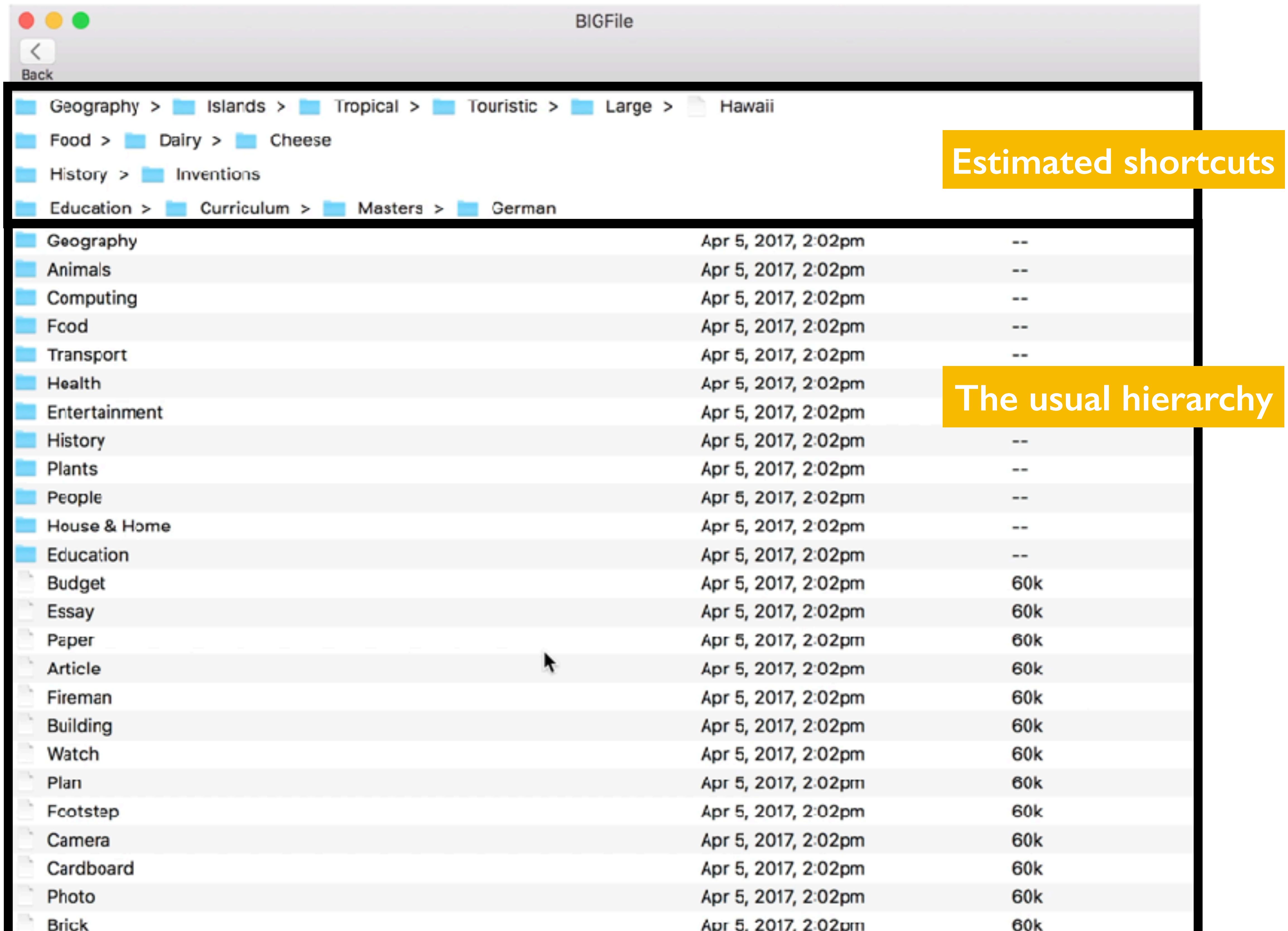
✻ **Leverages the expected information gain** $IG(\Theta|X = x, Y)$



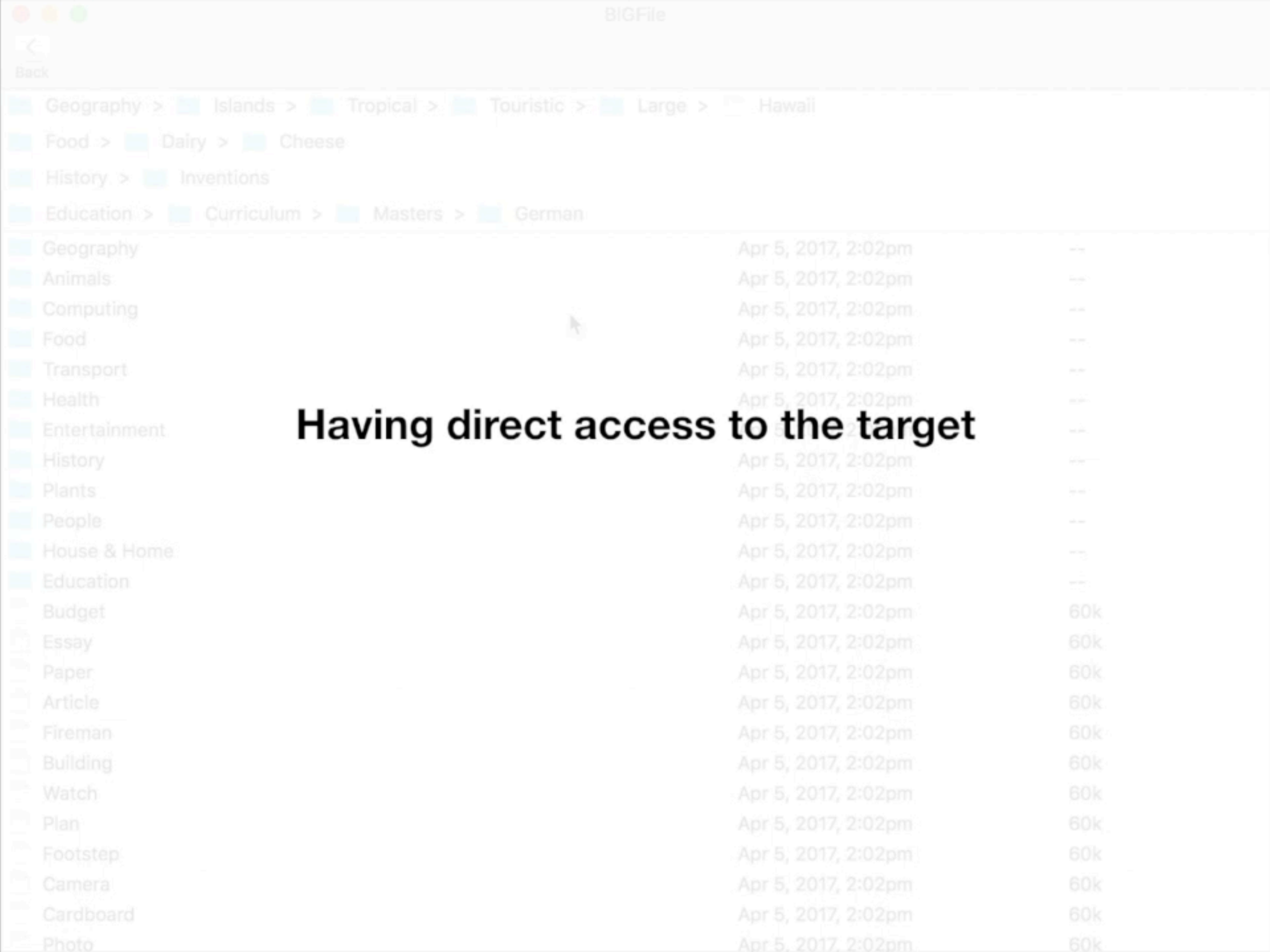
- Bayesian Information Gain (BIG) - BIGFile



- Bayesian Information Gain (BIG) - BIGFile



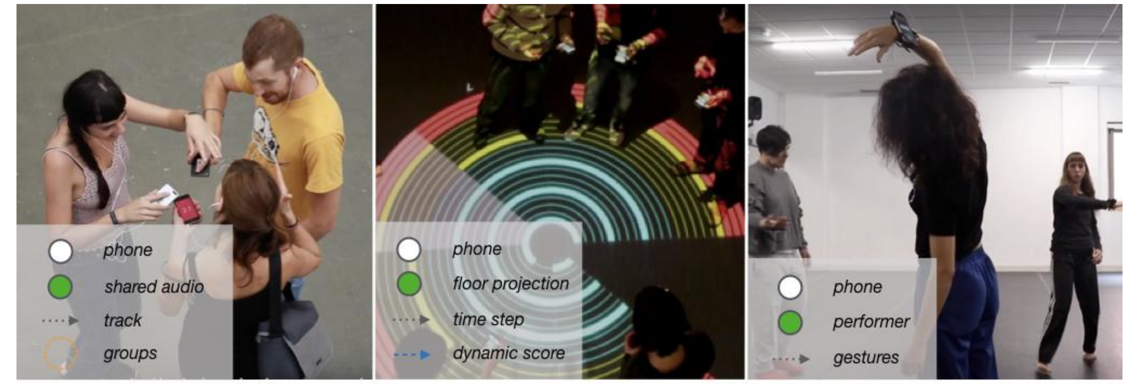
Having direct access to the target



- Bayesian Information Gain (BIG) - BIGnav, BIGFile

From goal-directed interaction
To computational creativity

- Collective Music Making



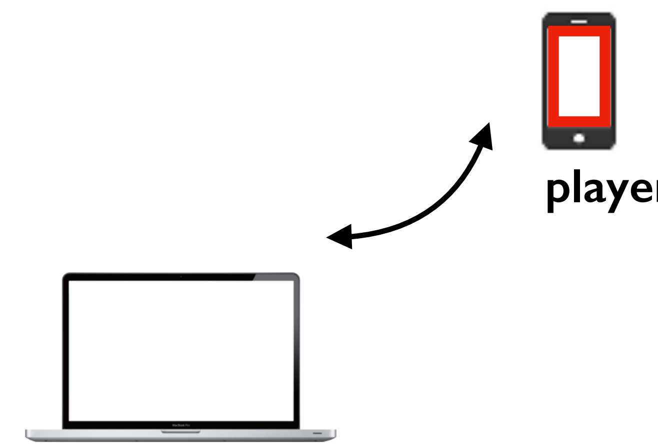
COSIMA

http://cosima.ircam.fr

coord. Norbert Schnell

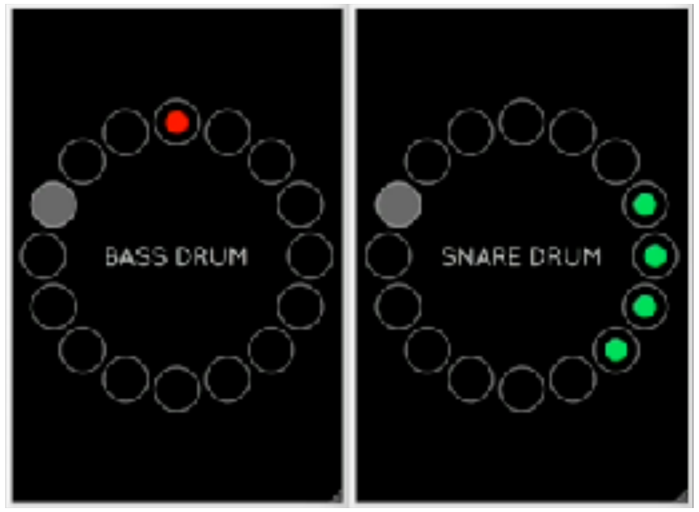
- Coloop (2017) - collective sequencer

- Coloop (2017) - collective sequencer

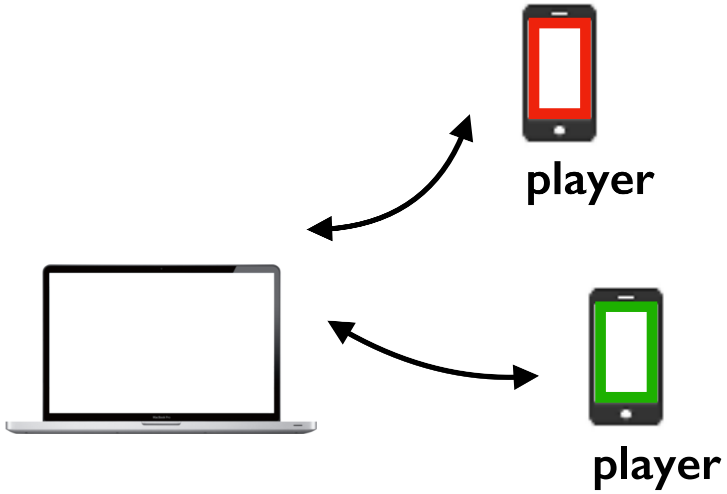


individual player interface

- Coloop (2017) - collective sequencer



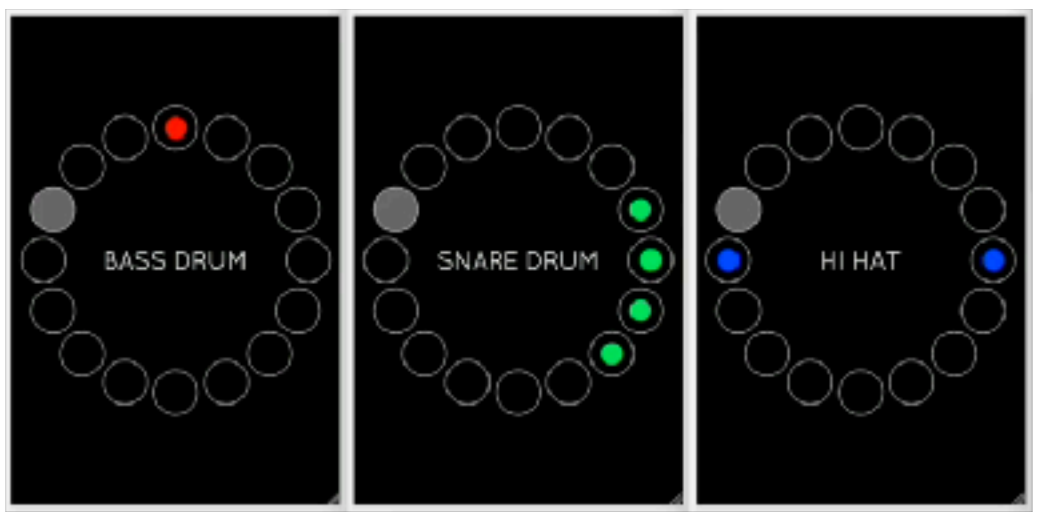
individual player interface



Synchronized clock

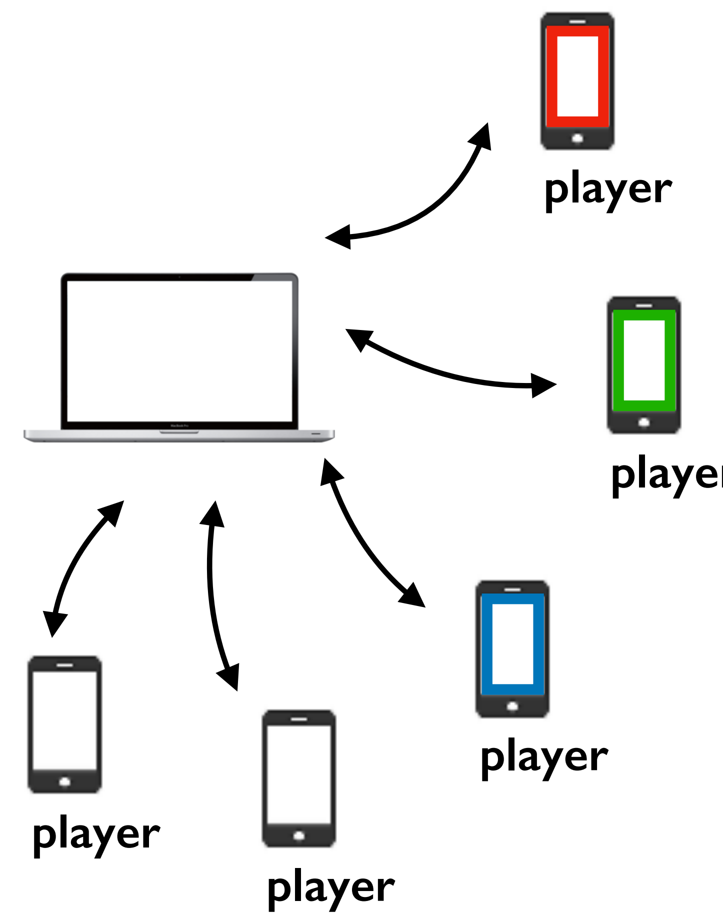
https://github.com/collective-soundworks

- Coloop (2017) - collective sequencer



individual player interface

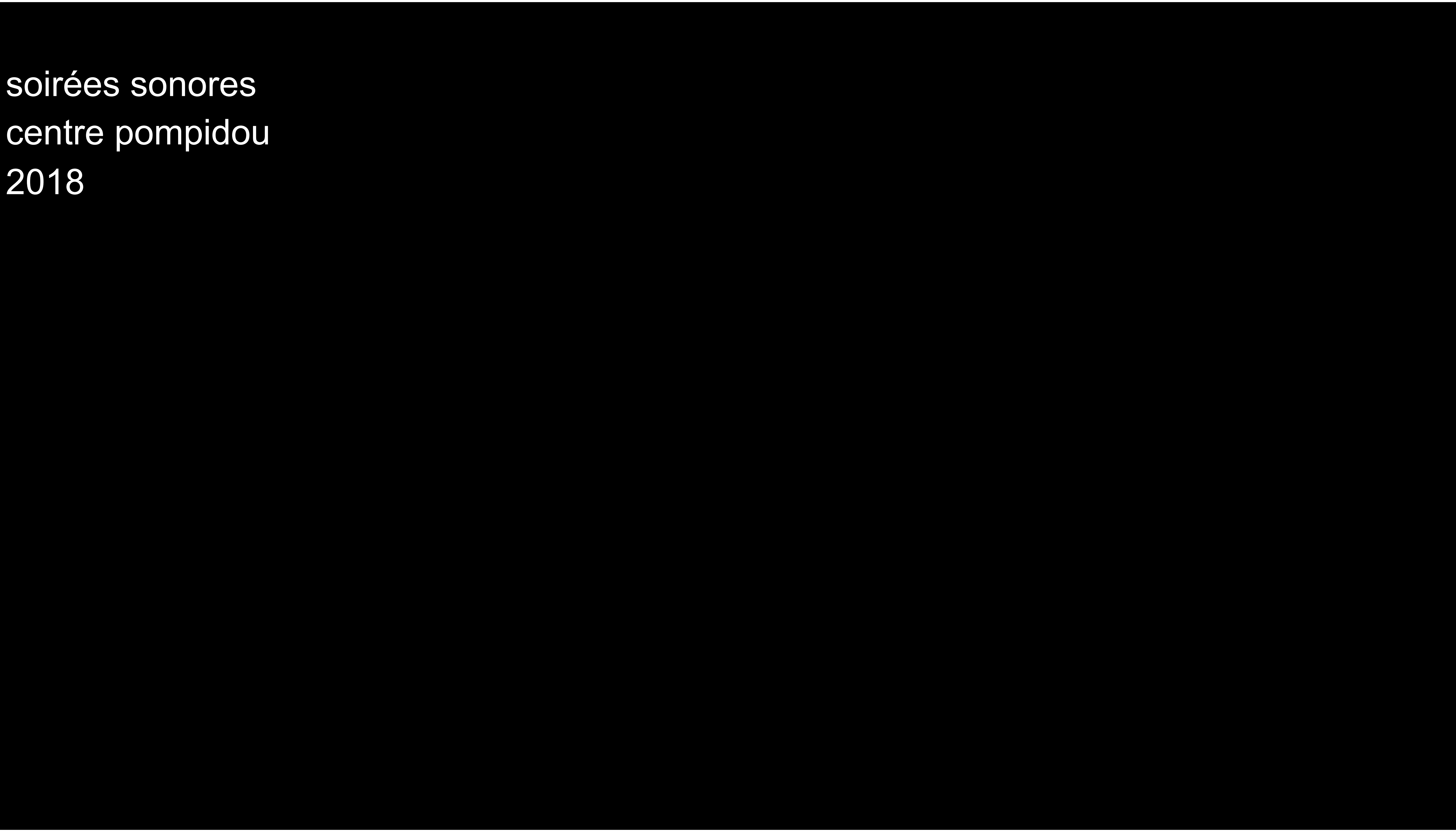
Shared music object



Synchronized clock

https://github.com/collective-soundworks

- Coloop (2017) - collective sequencer



- Coloop (2017) - collective sequencer

musical outcome is **collective**…

how to encourage social interaction?

…but most participants remained **individual**

- Coloop (2017) - collective sequencer

- Entrain (2019) - adaptive musical agent

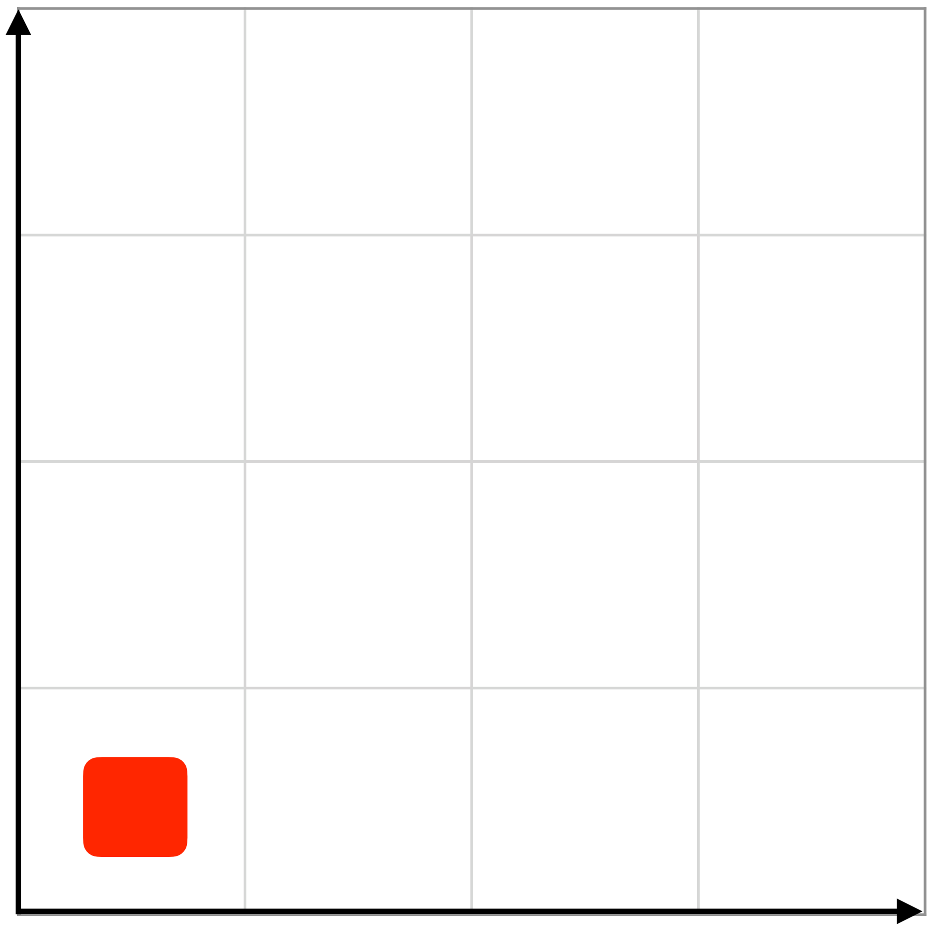
- Entrain (2019) - adaptive musical agent

Periodicity

Activity

Player features

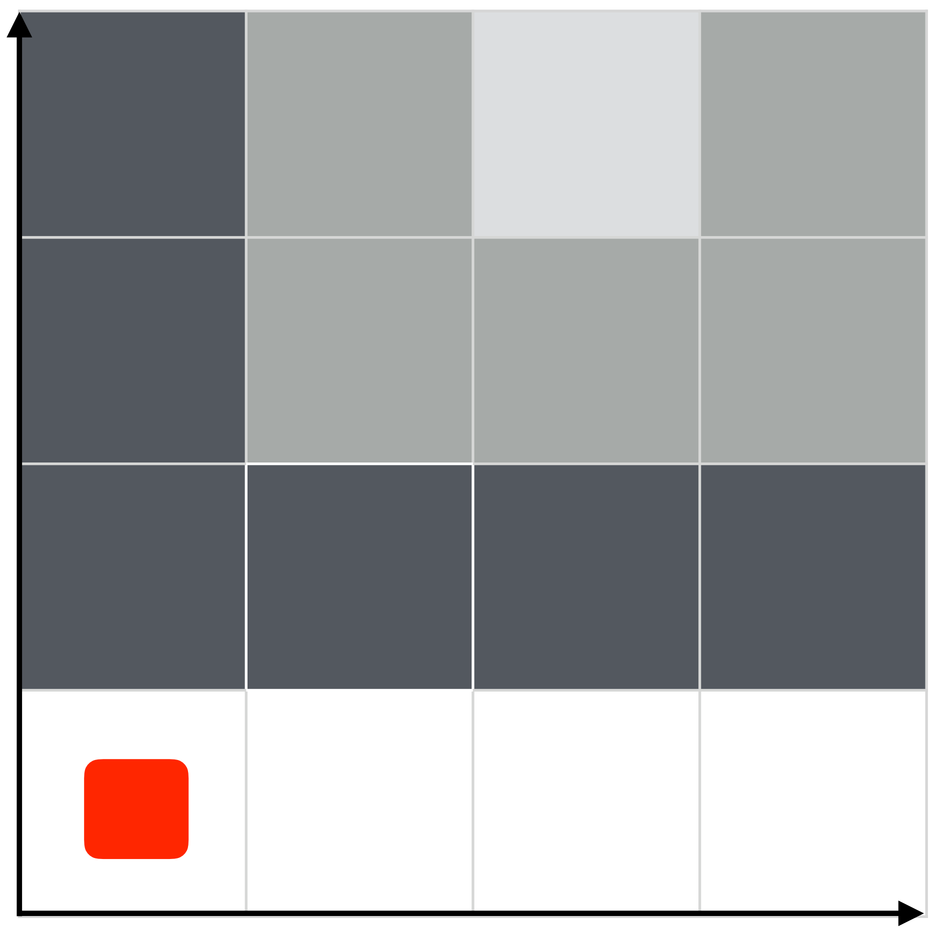
- Entrain (2019) - adaptive musical agent



Player features

$Y = y$ Player state

- Entrain (2019) - adaptive musical agent

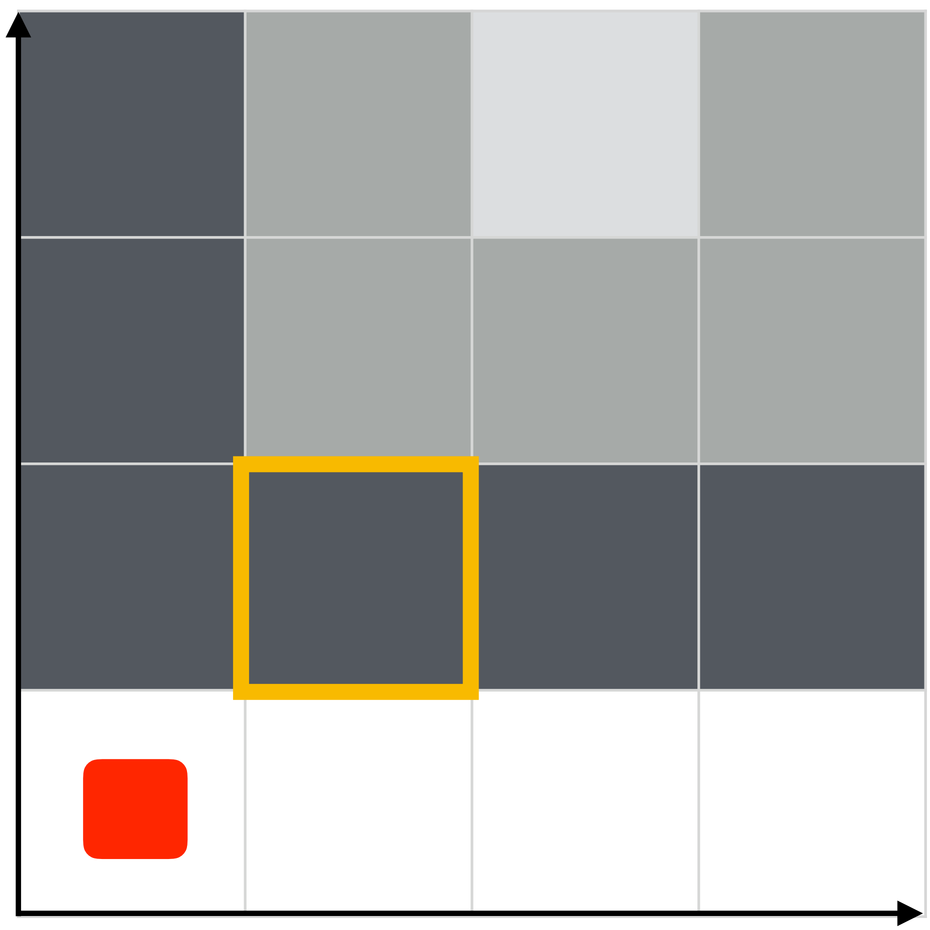


Player features

$P(\Theta = \theta)$ Probabilistic model

$Y = y$ Player state

- Entrain (2019) - adaptive musical agent



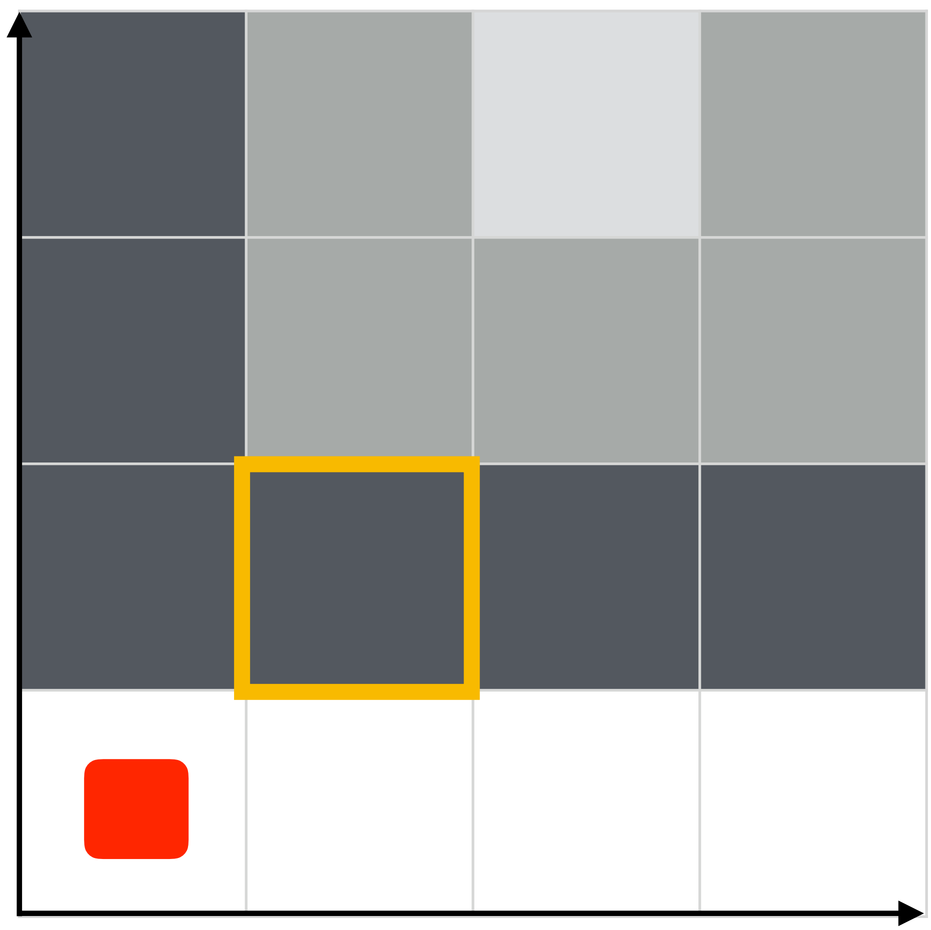
Player features

$P(\Theta = \theta)$ Probabilistic model

$X = x$ Special state

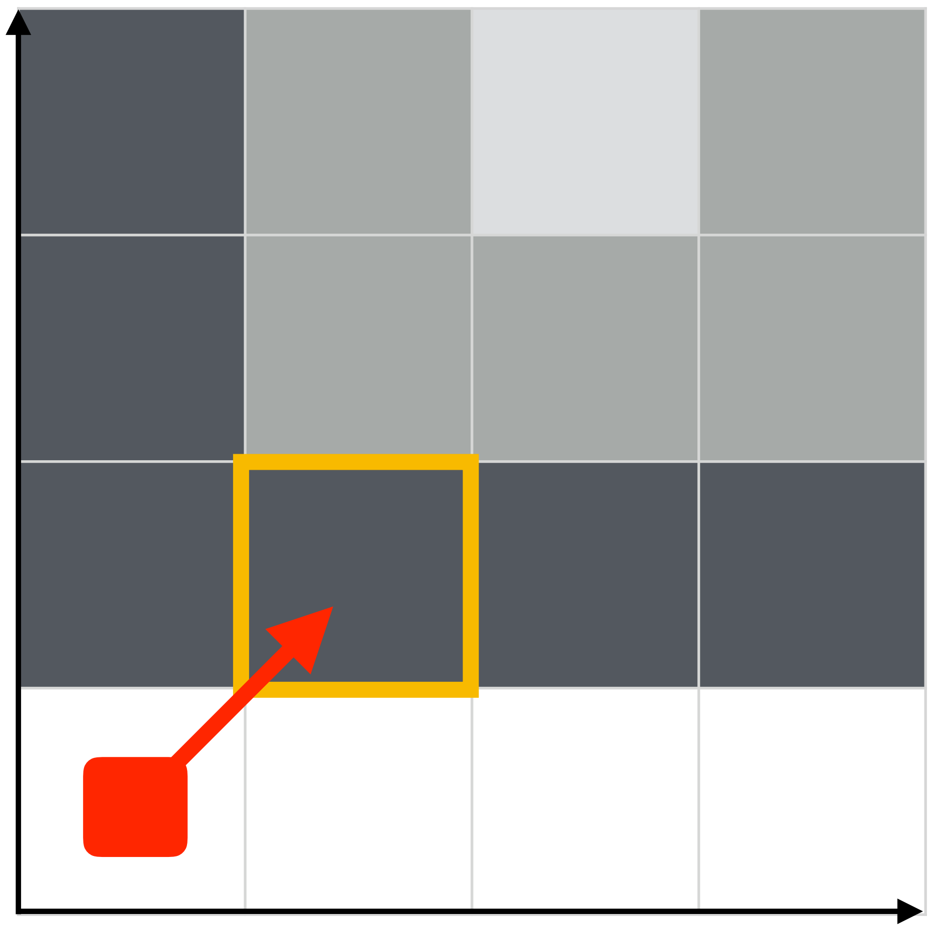
$Y = y$ Player state

- Entrain (2019) - adaptive musical agent



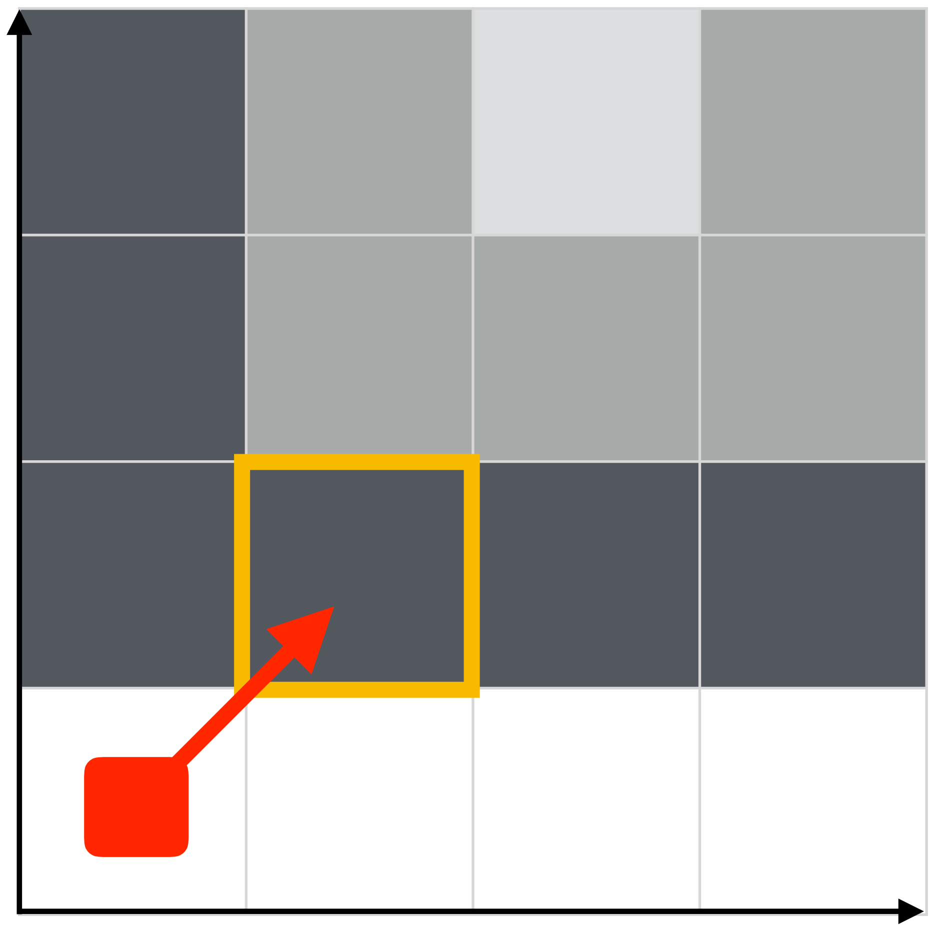
Player features

- Entrain (2019) - adaptive musical agent



Player features

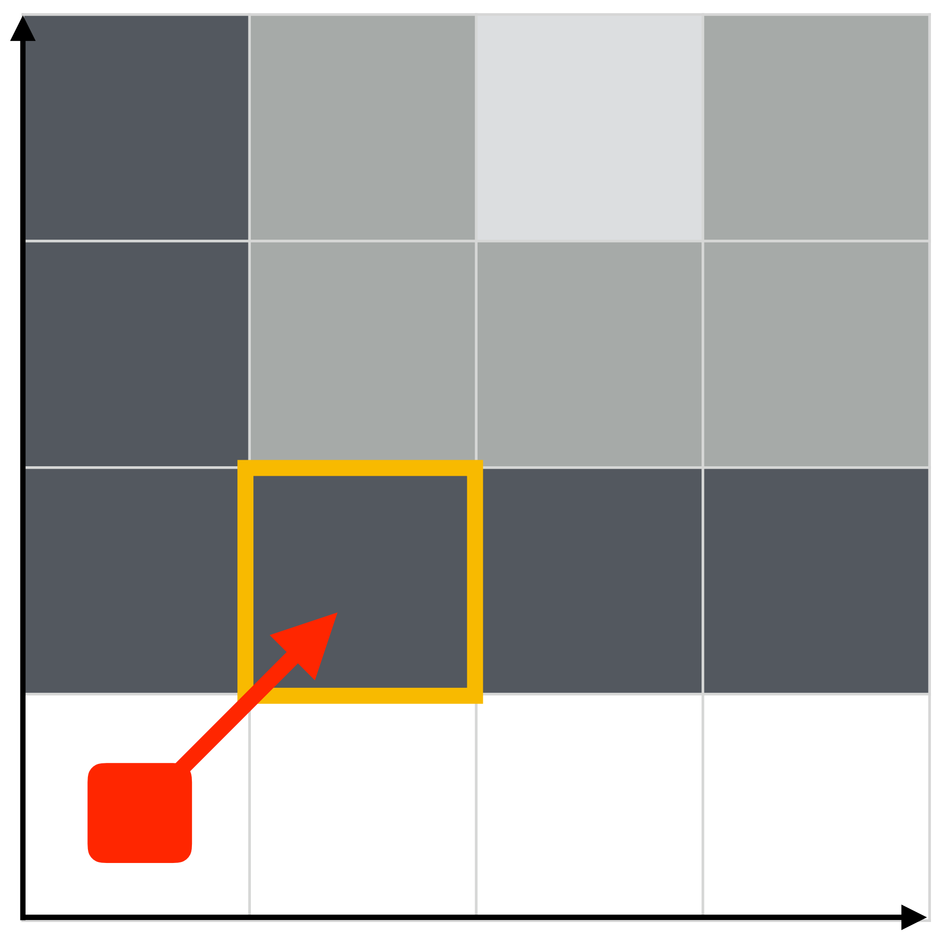
- Entrain (2019) - adaptive musical agent



Player features

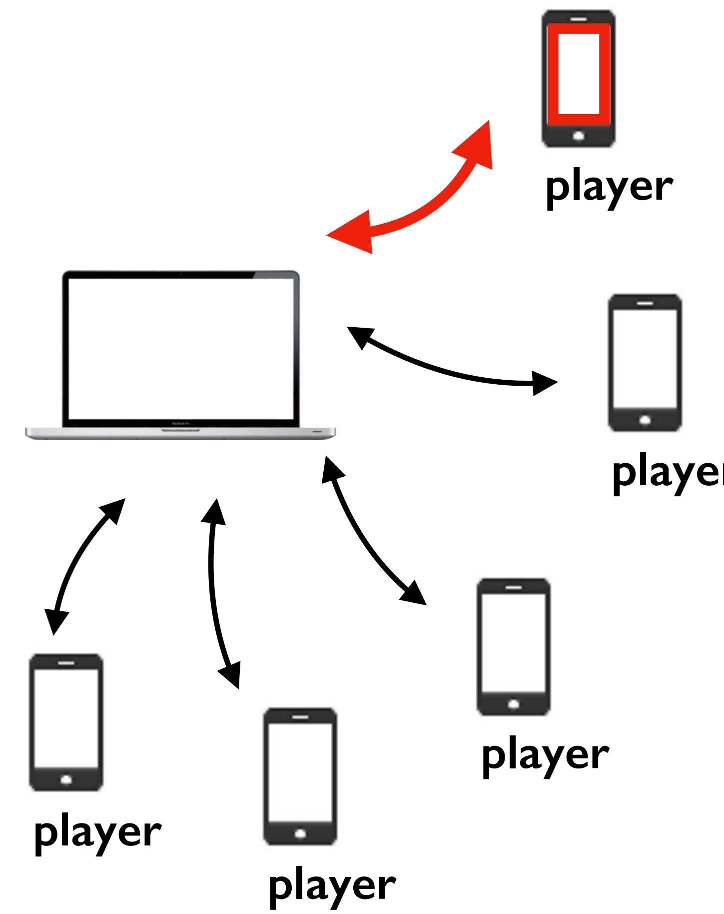
Shared visual effect

- Entrain (2019) - adaptive musical agent



Player features

Shared visual effect



Shared audio effect

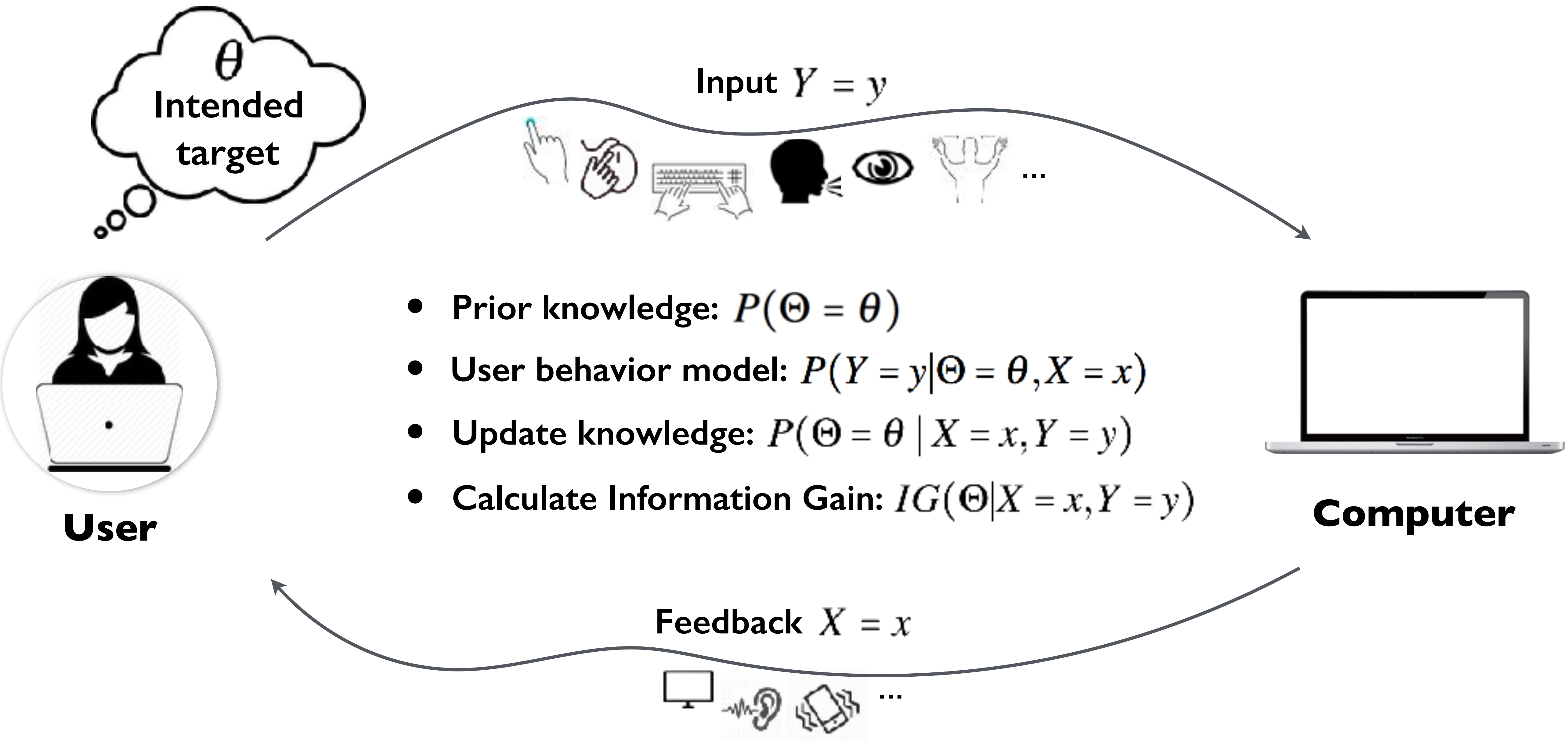
- Entrain @ SIGGRAPH'19

- Entrain @ Computational Interaction Summer School

Connect yourself to

192.168.1.100:8000

- 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, Entrain**



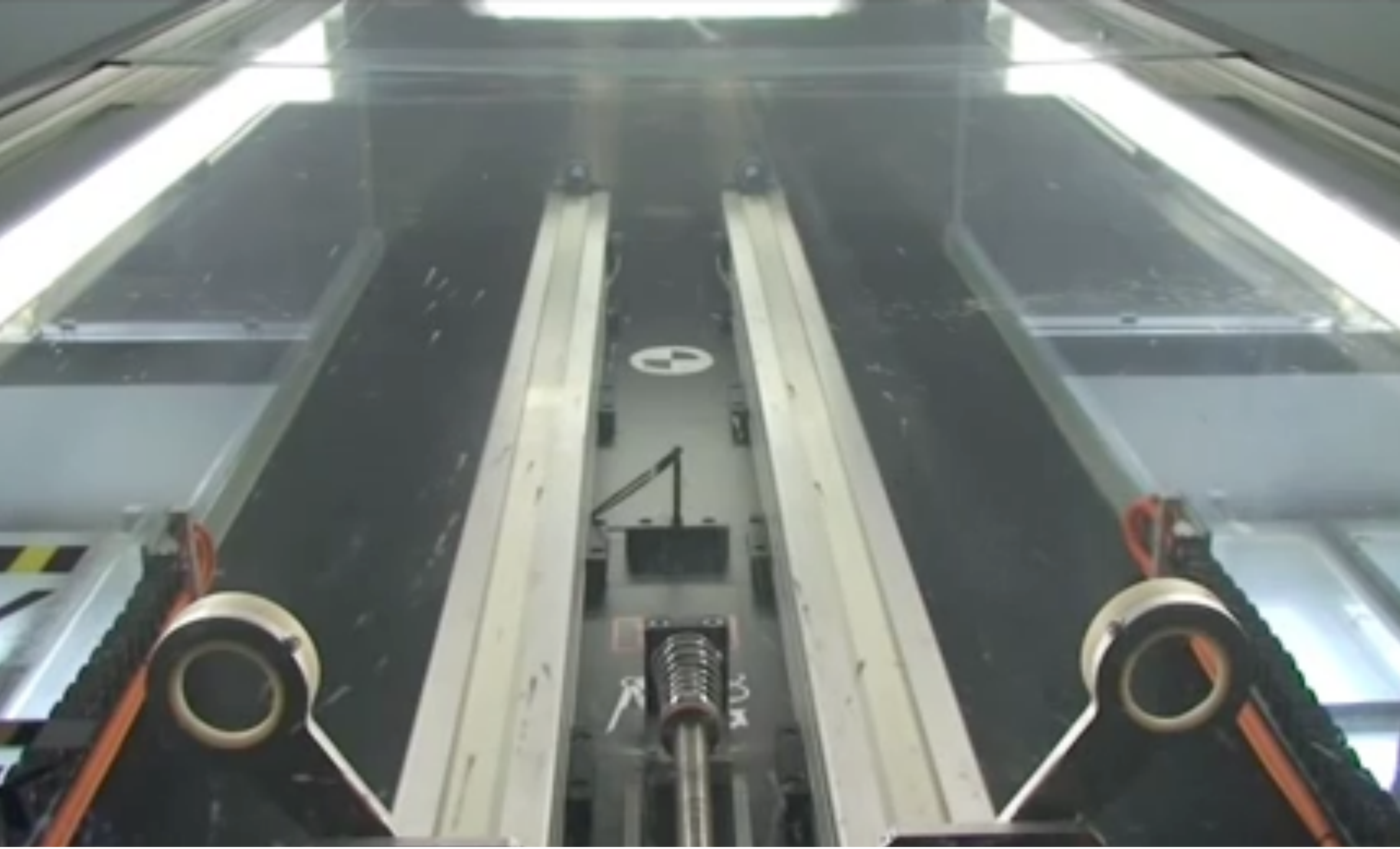
- Prior knowledge: $P(\Theta = \theta)$
- User behavior model: $P(Y = y|\Theta = \theta, X = x)$
- Update knowledge: $P(\Theta = \theta \mid X = x, Y = y)$
- Calculate Information Gain: $IG(\Theta|X = x, Y = y)$

Recap

* Information can be quantified in bits
* Information reduces uncertainty
* Improved information gain (BIG)
* Information-theoretic measures to characterize interaction

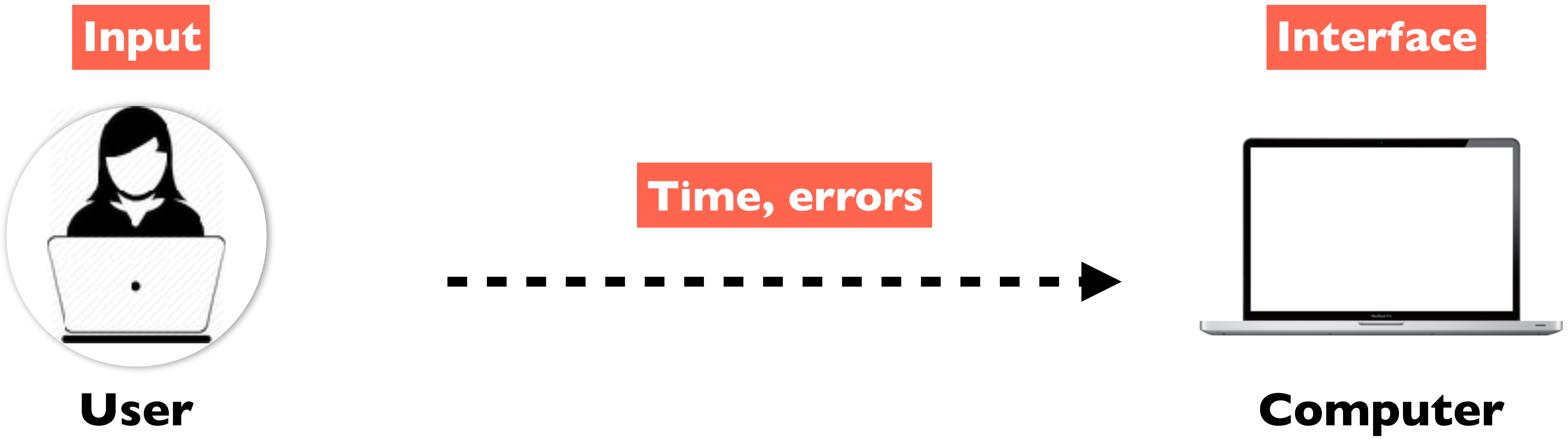
Email: AbbyWanyu.Liu@ircam.fr
Web: abbywanyuliu.com

Email: AbbyWanyu.Liu@ircam.fr
Web: abbywanyuliu.com



Email: AbbyWanyu.Liu@ircam.fr
Web: abbywanyuliu.com

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction

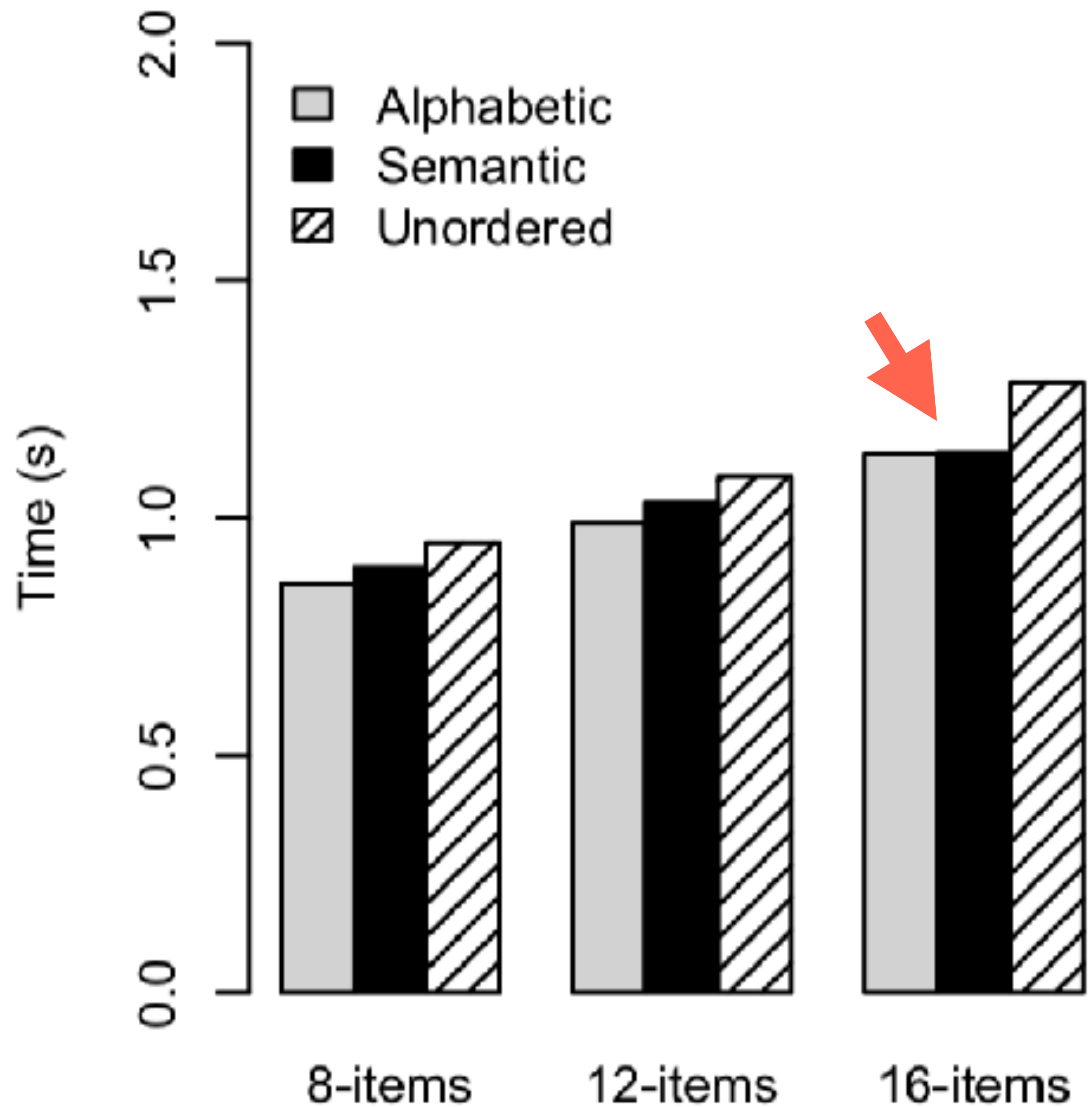
Drawback 1: Speed-accuracy tradeoff

- 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)

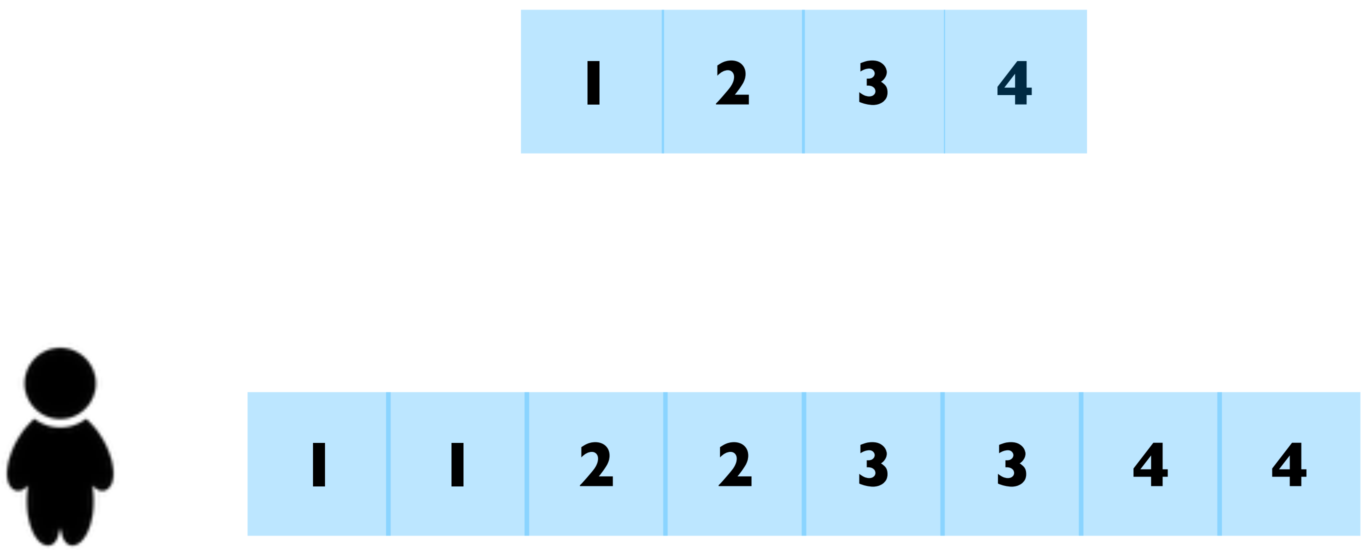
- Information-theoretic measures to characterize interaction

Drawback 2: The treatment of errors

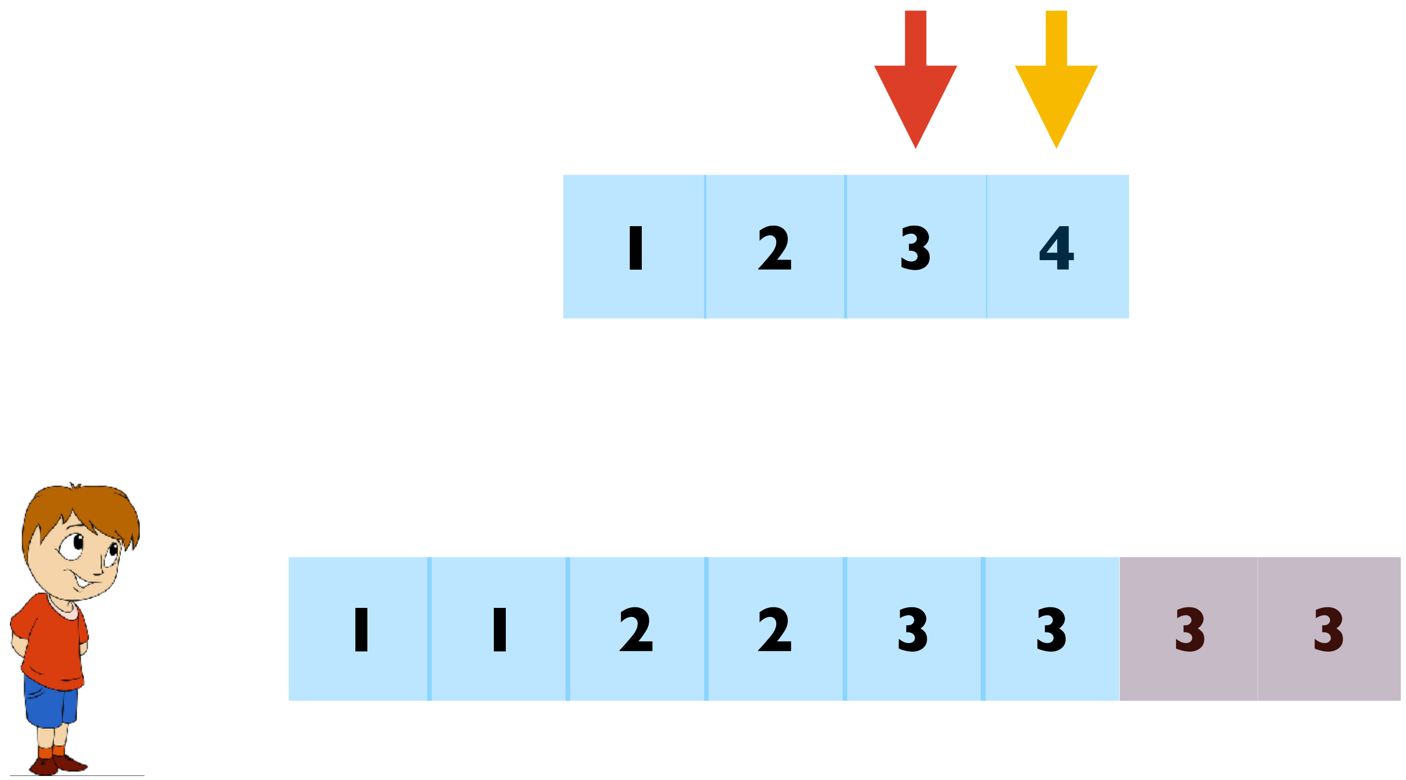
- Information-theoretic measures to characterize interaction

1	**2**	**3**	**4**

- Information-theoretic measures to characterize interaction

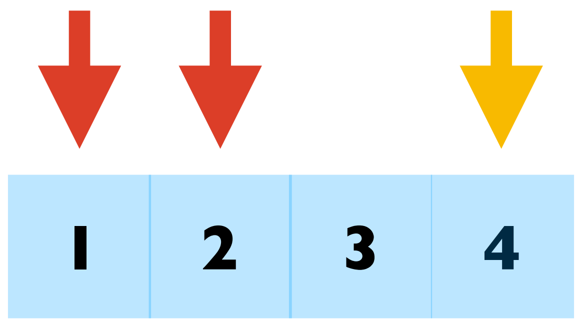


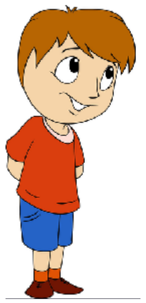
- Information-theoretic measures to characterize interaction



Error rate: 2 / 8 = 25 %

- Information-theoretic measures to characterize interaction





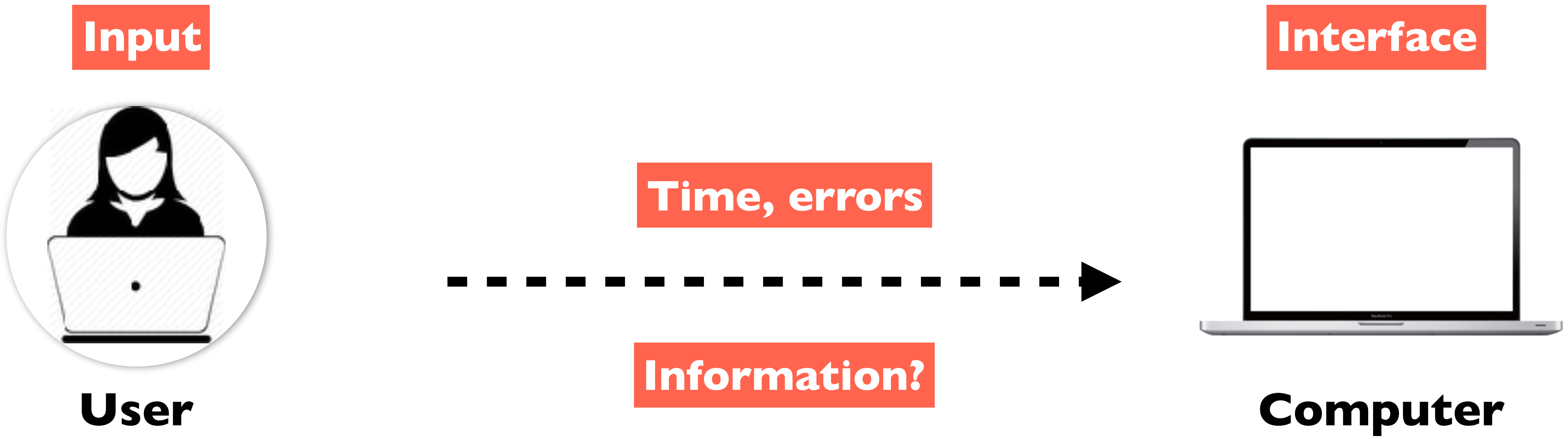
1	1	2	2	3	3	3	3

Error rate: 2 / 8 = 25 %

1	1	2	2	3	3	1	2

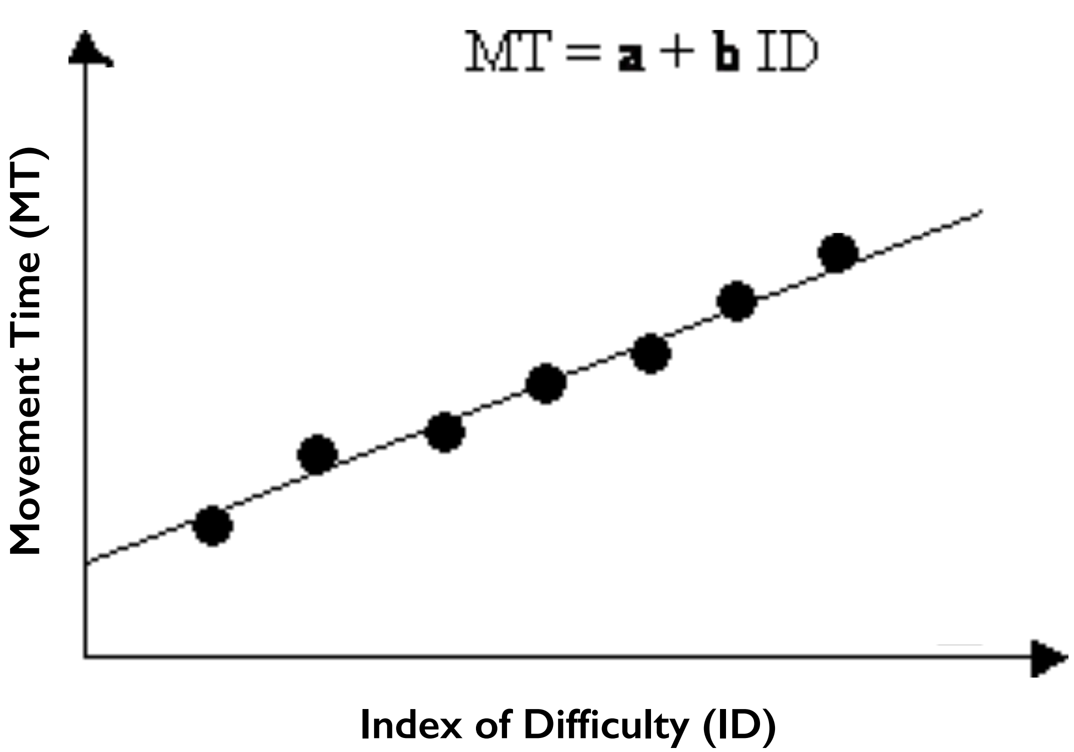
Error rate: 2 / 8 = 25 %

- Information-theoretic measures to characterize interaction



- Information-theoretic measures to characterize interaction

Fitts' law

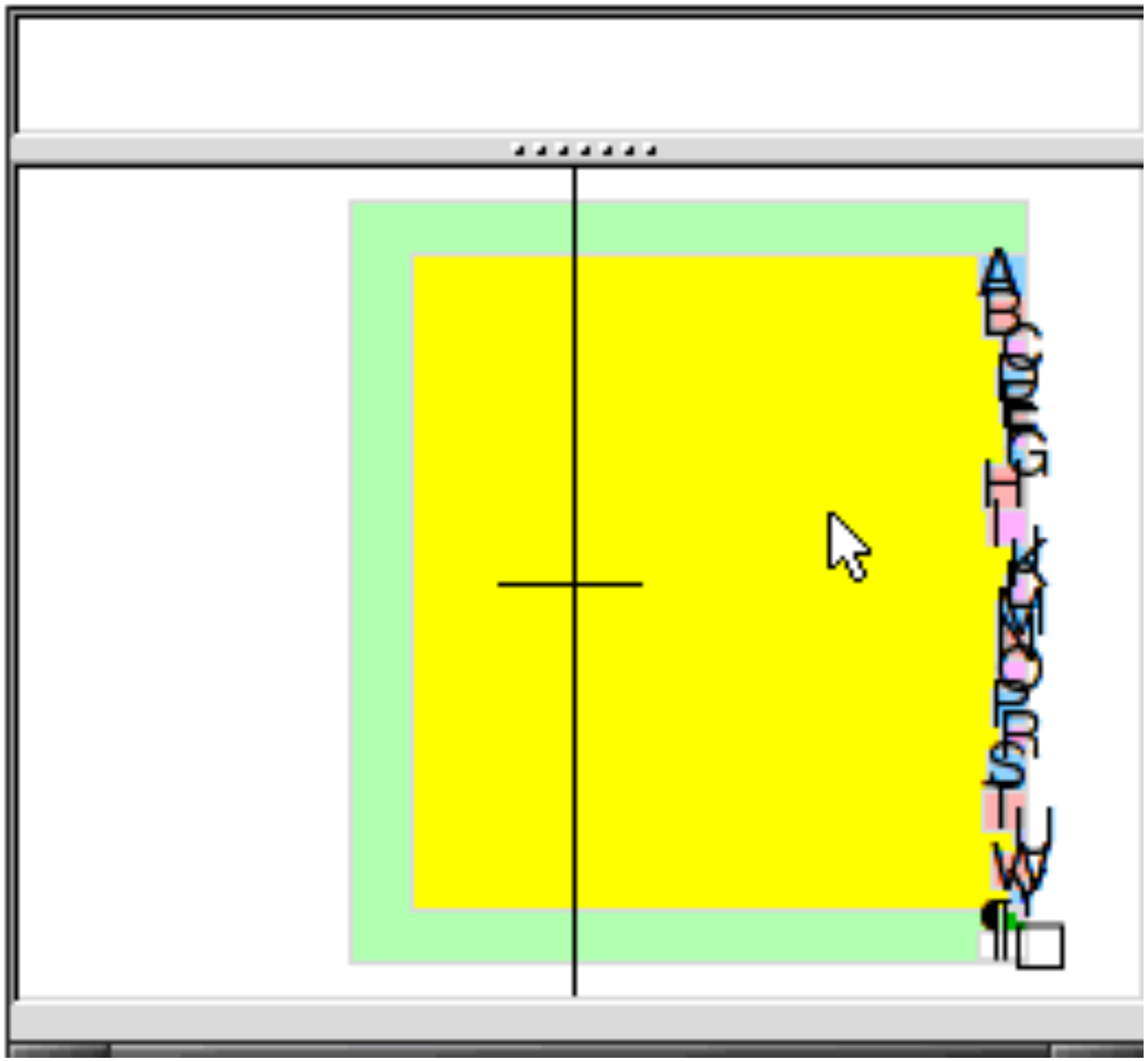


Movement Time: $T = a + b \log_2\left(1 + \frac{D}{W}\right)$

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

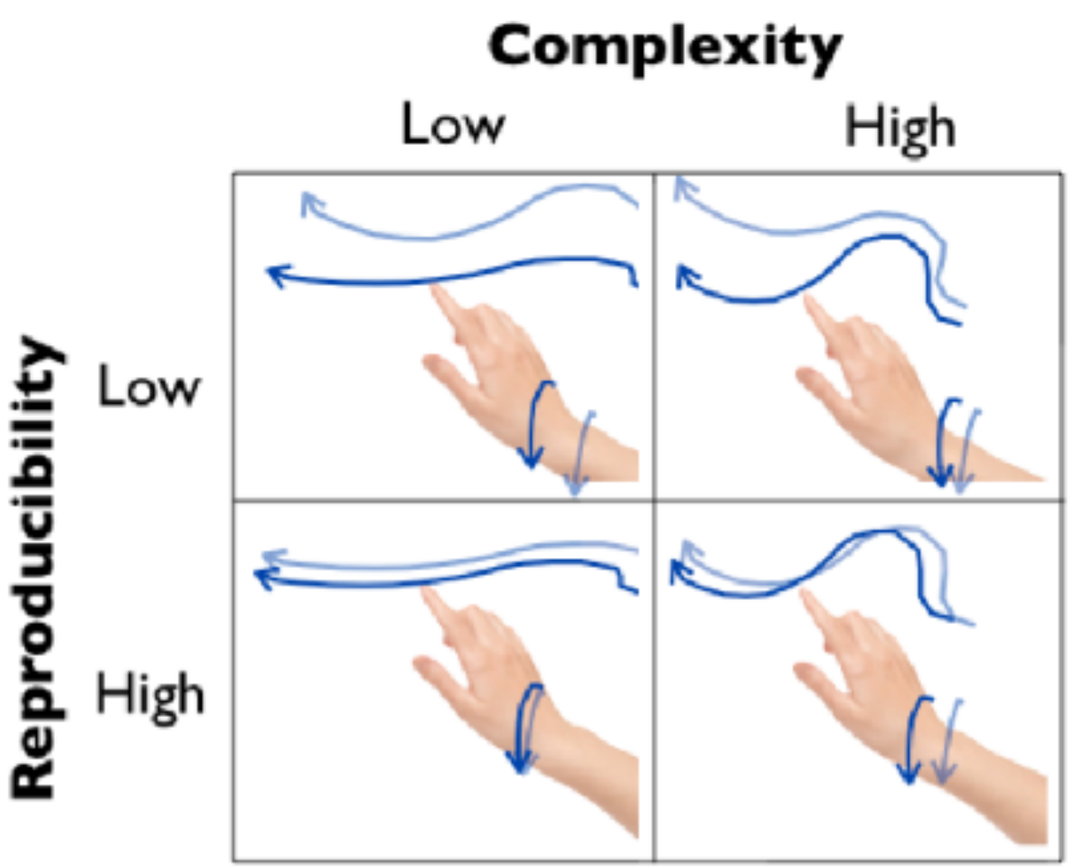
- Information-theoretic measures to characterize interaction

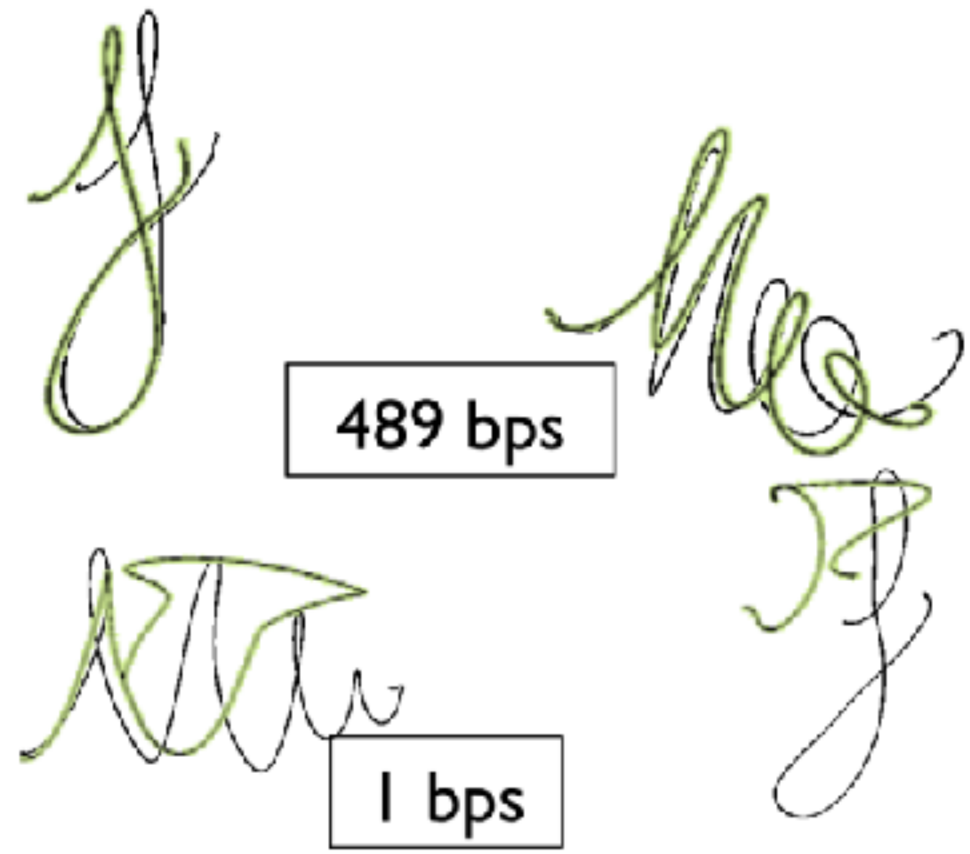
Dasher



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

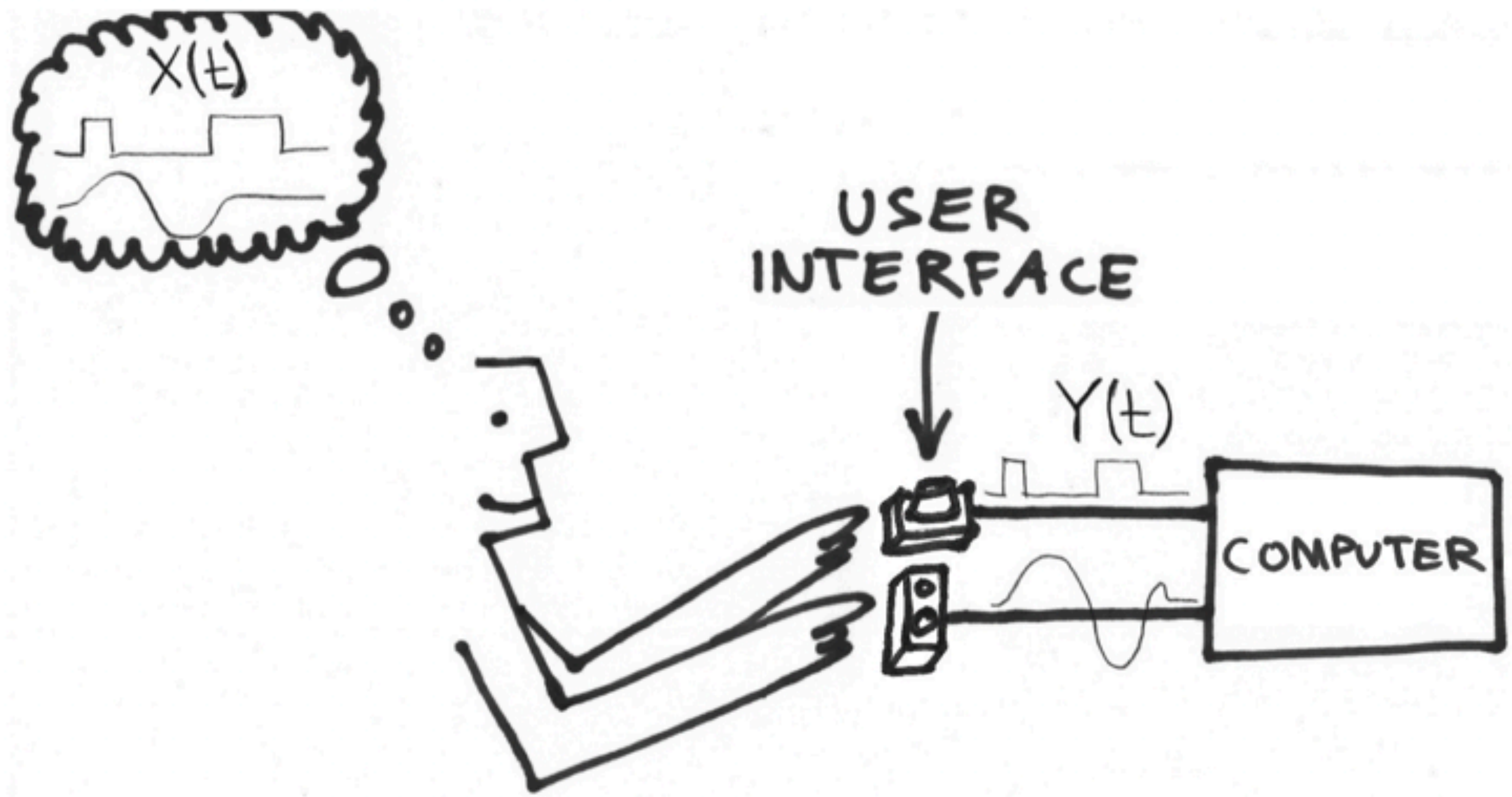
- Information-theoretic measures to characterize interaction





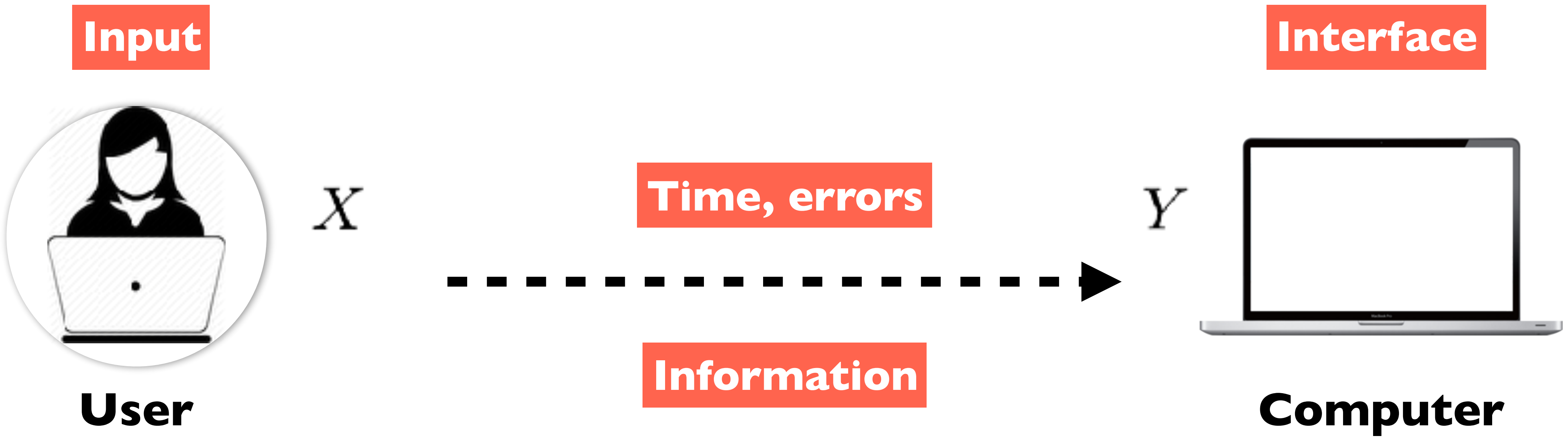
Information Capacity of Full-body Movements. (Oulasvirta et al. CHI'13)

- 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 takes values in

1	**2**	**3**	**4**
x_1	x_2	x_3	x_4

- Information-theoretic measures to characterize interaction

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

X takes values in

1	2	3	4
$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$

- Information-theoretic measures to characterize interaction

Input entropy: How much information could be transmitted.

Corresponding to **input size** and the **probability distribution**.

1	**2**	**3**	**4**

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

- Information-theoretic measures to characterize interaction

Y: The actual input received by the computer.

1	**2**	**3**	**4**

- Information-theoretic measures to characterize interaction

Y: The actual input received by the computer.

1	2	3	4

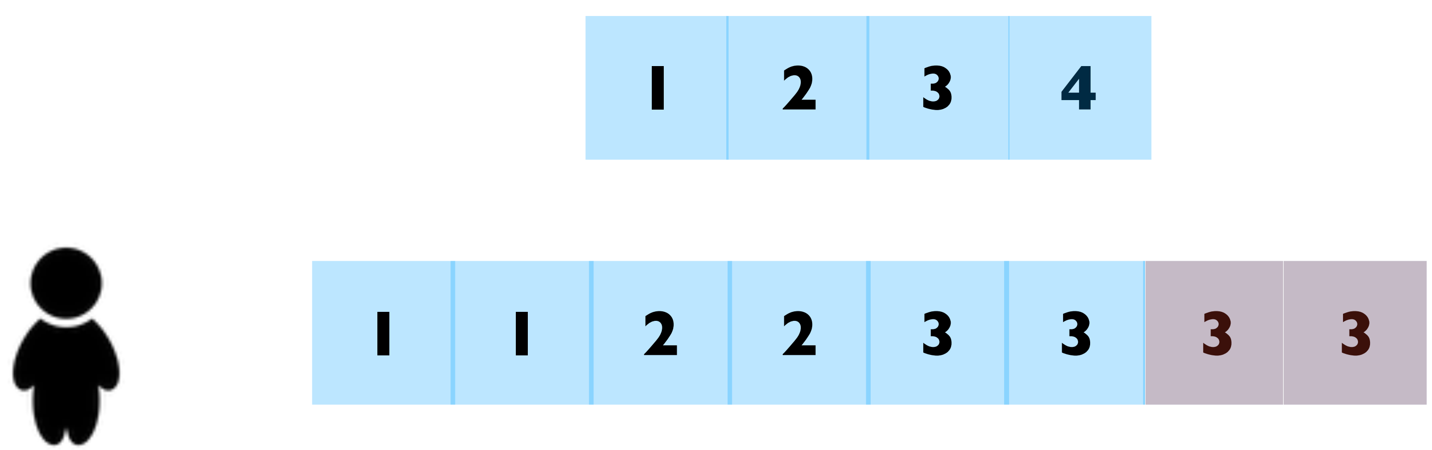
1	1	2	2	3	3	4	4

$H(X) = 2$ bits

- 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.

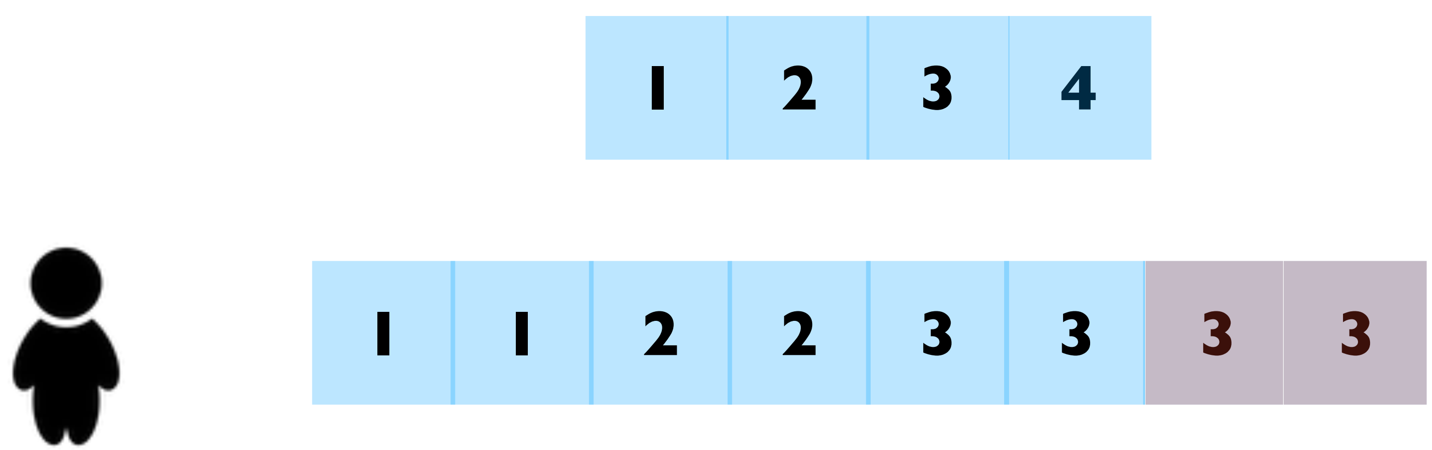


$$I(X;Y) = H(X) - H(X|Y)$$

- 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.

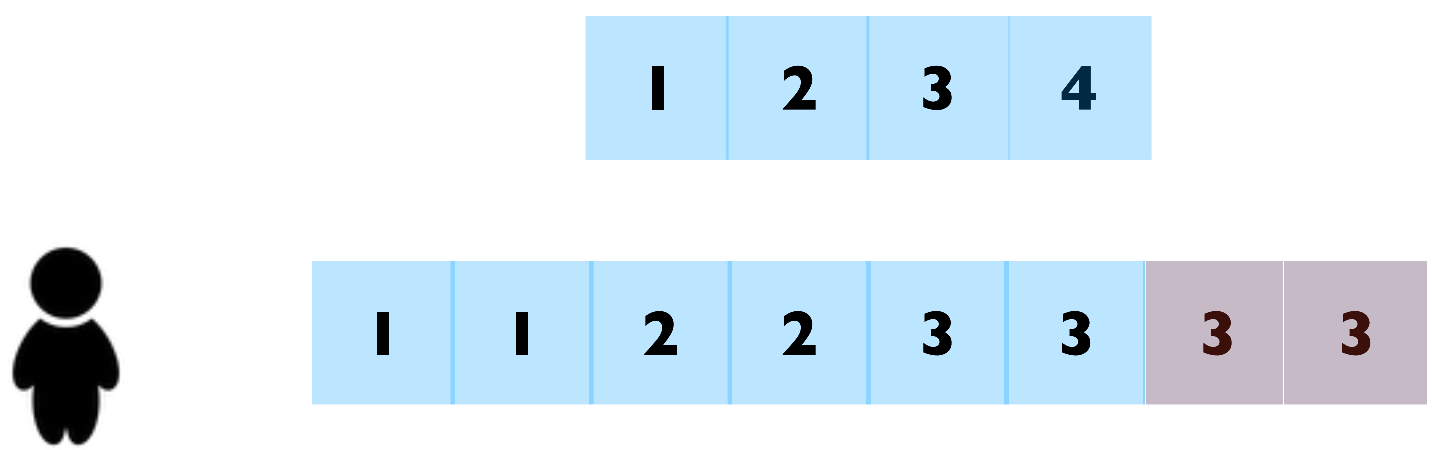


$$I(X;Y) = \sum_x \sum_y P(X=x, Y=y) \log_2 \frac{P(X=x, Y=y)}{P(X=x)P(Y=y)}$$

- Information-theoretic measures to characterize interaction

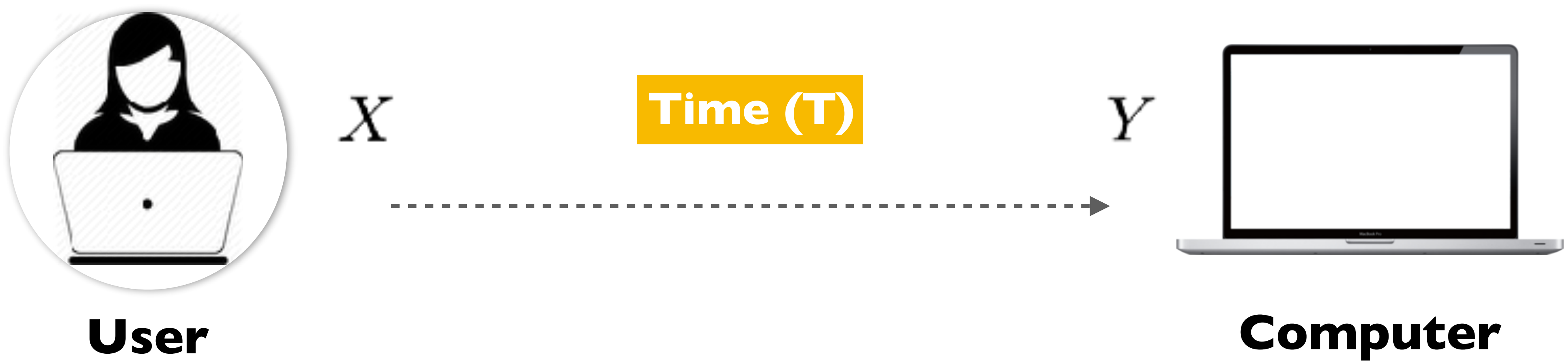
$I(X;Y)$: Mutual information between the intended input and the actual input.

It describes how much information actually gets transmitted.



$I(X;Y) = 1.5$ bits

- Information-theoretic measures to characterize interaction



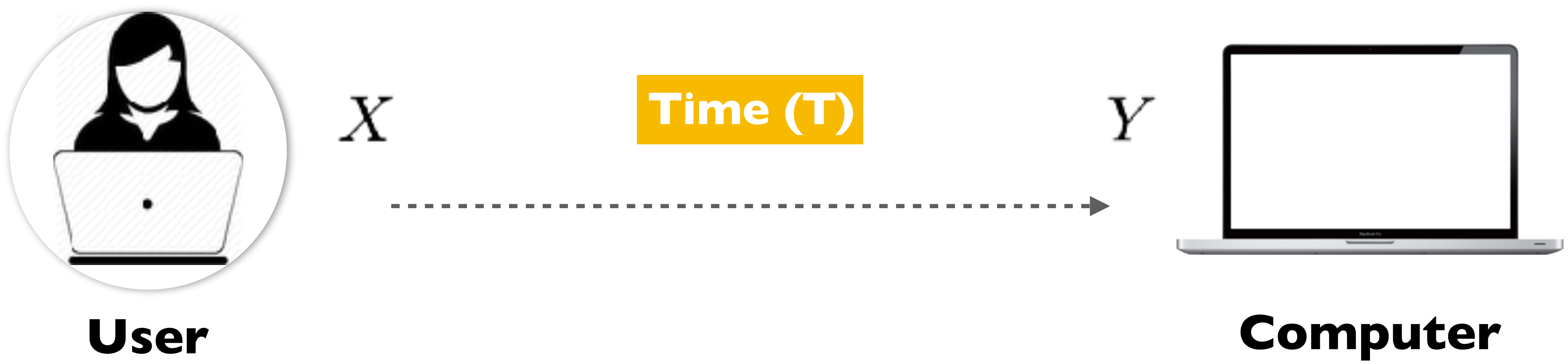
$$TP = \frac{I(X;Y)}{T}$$:Throughput describes information transmission efficiency.

1	2	3	4

1	1	2	2	3	3	3	3

$TP = I(X;Y)/T = 1.5/1.5 = 1$ bit / s

- Information-theoretic measures to characterize interaction



$$TP = \frac{I(X;Y)}{T}$$:Throughput describes information transmission efficiency.

1	2	3	4

1	1	2	2	3	3	4	4

$$TP = I(X;Y)/T = H(X)/T = 2/1.5 = 1.3 \text{ bit / s}$$

- 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:

More consistent measure

- Information-theoretic measures to characterize interaction

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP_i
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

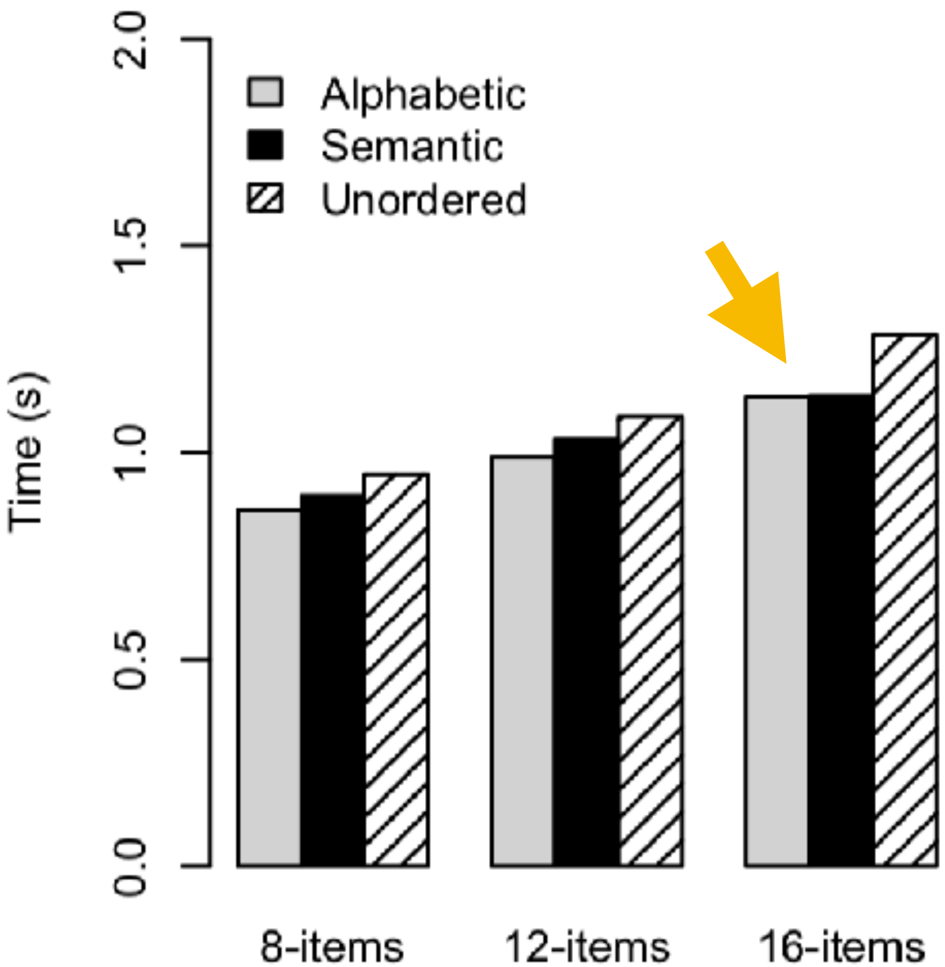
Advantage 3:

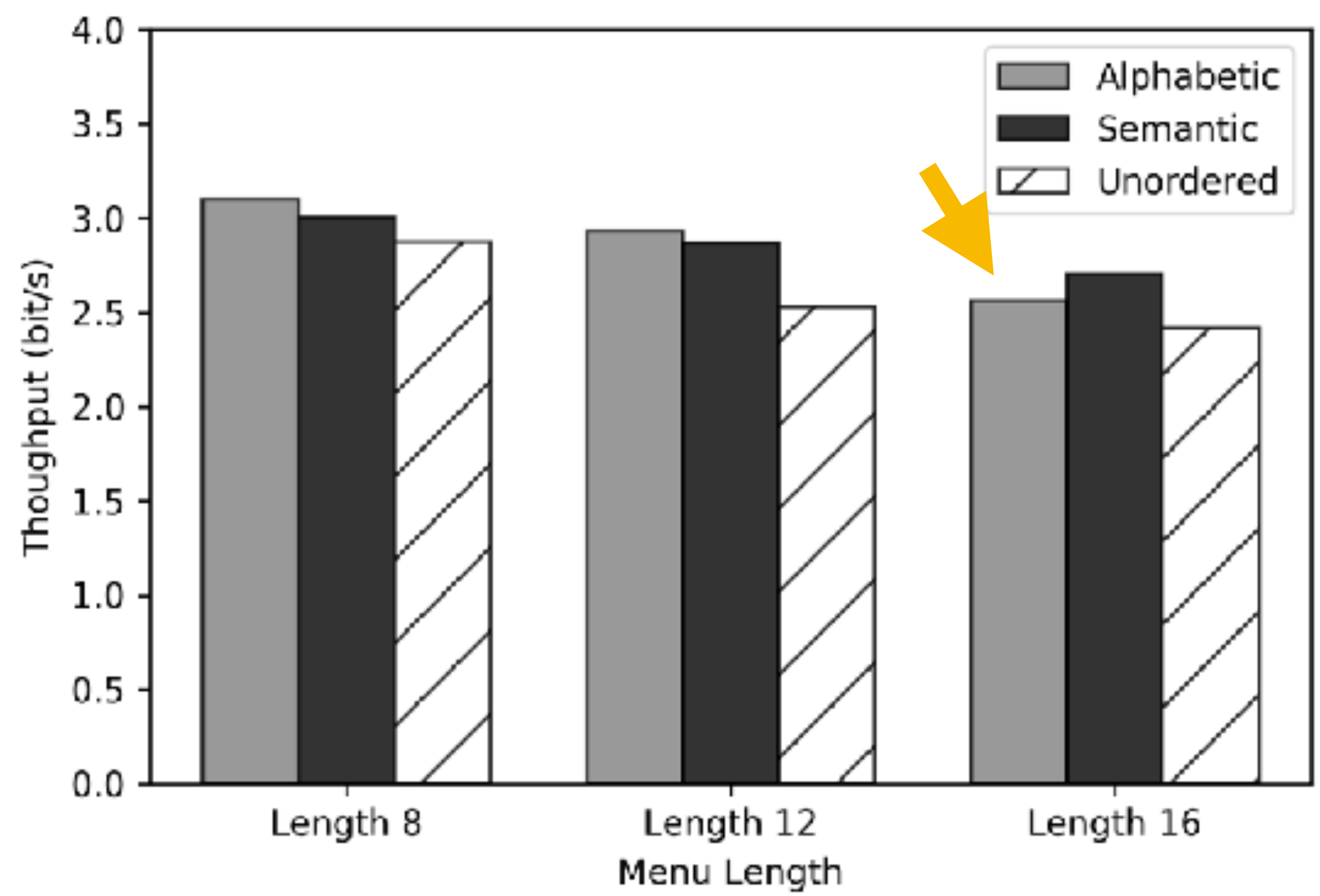
Speed-accuracy tradeoff

- Information-theoretic measures to characterize interaction

Advantage 3:

Speed-accuracy tradeoff





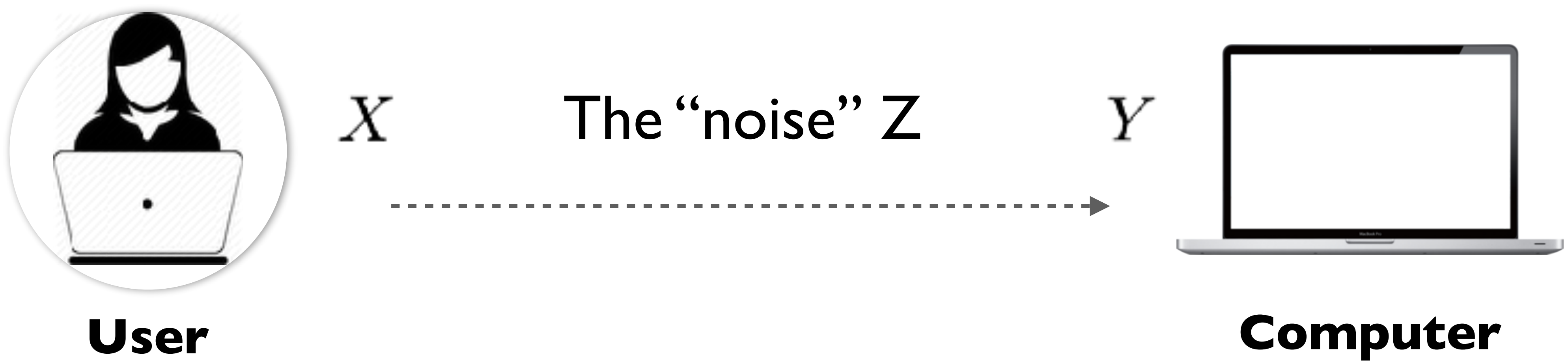
- Information-theoretic measures to characterize interaction

Advantage 4:

Equivocation provides information about

how users make errors

- Information-theoretic measures to characterize interaction



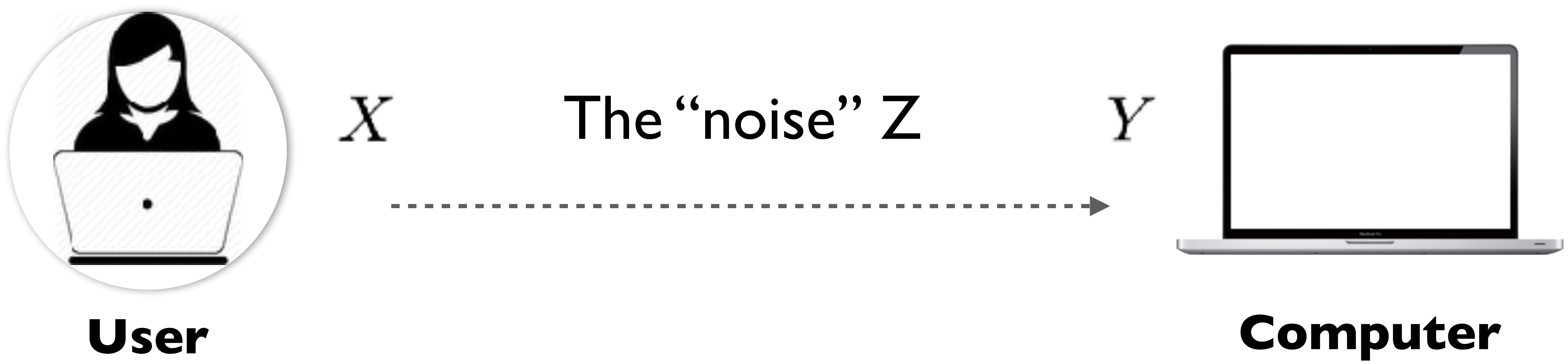
The error random variable:

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

The error rate:

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

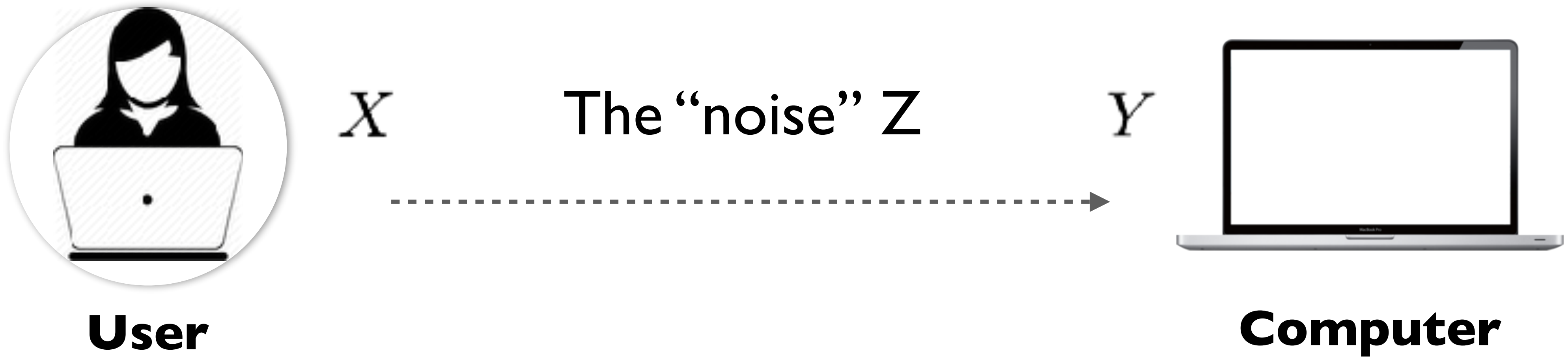
- Information-theoretic measures to characterize interaction



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



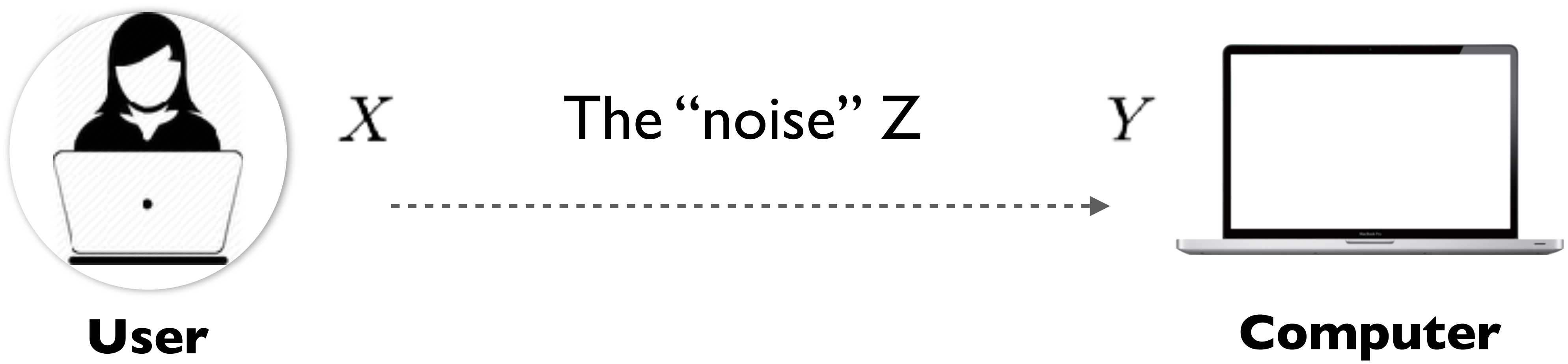
$$I(X;Y) = H(X) - H(X|Y)$$

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

Fano's inequality.

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

- Information-theoretic measures to characterize interaction

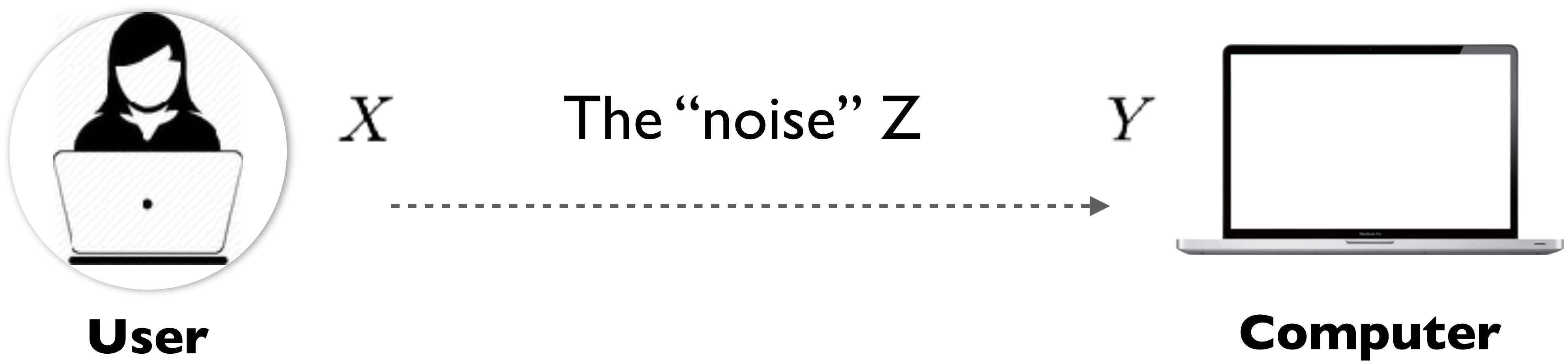


$$I(X;Y) = H(X) - H(X|Y)$$

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

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

- Information-theoretic measures to characterize interaction



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

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

How they make errors.

- Information-theoretic measures to characterize interaction

1	2	3	4

1	1	2	2	3	3	3	3

$H(X|Y) = 0.5$ bits

$TP = I(X;Y)/T$

$= 1.5/1.5 = 1$ bit / s

Error rate: 2 / 8 = 25 %

1	1	2	2	3	3	1	2

$H(X|Y) = 0.7$ bits

$TP = I(X;Y)/T$

$= 1.3/1.5 = 0.8$ bit / s

Error rate: 2 / 8 = 25 %

Exercise