PhD Dissertation Defense at Telecom ParisTech on 22.11.2018

Information Theory as a Unified Tool for Understanding and Designing Human-Computer Interaction

Abby Wanyu Liu

Committee:

#HumanComputerInteraction #MutualInformation #Bayesian #ComputationalInteraction #CoAdaptation

Because human-computer interaction studies

a human

ACM SIGCHI Curriculum for human-computer interaction (2009)

Because human-computer interaction studies

a human

User

ACM SIGCHI Curriculum for human-computer interaction (2009)

Because human-computer interaction studies

a human and a machine

ACM SIGCHI Curriculum for human-computer interaction (2009)

Because human-computer interaction studies a human and a machine

User

Computer

ACM SIGCHI Curriculum for human-computer interaction (2009)

Because human-computer interaction studies
a human and a machine in **communication**,

User

Computer

ACM SIGCHI Curriculum for human-computer interaction (2009)

Because human-computer interaction studies
a human and a machine in **communication**,

User

communication

Computer

ACM SIGCHI Curriculum for human-computer interaction (2009)

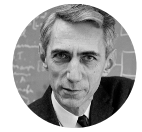
Because human-computer interaction studies a human and a machine in **communication**, it draws from supporting knowledge on both the machine and the human side.

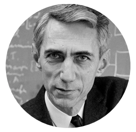
User

communication

Computer

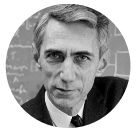
ACM SIGCHI Curriculum for human-computer interaction (2009)

1948 **Information Theory** 

1948 **Information Theory** 

1950 **Experimental Psychology Applications**

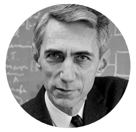
1948 Information Theory



1950 Experimental Psychology Applications

Hick's law (1952)

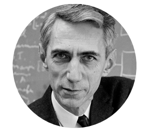
1948 **Information Theory**



1950 **Experimental Psychology Applications**

- **Hick's law** (1952)

- **Fitts' law** (1954)

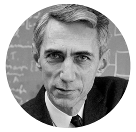
1948 **Information Theory** 

1950 **Experimental Psychology Applications**

- **Hick's law** (1952) 
- **Fitts' law** (1954) 

2017 **Bayesian Information Gain**

1948 **Information Theory**

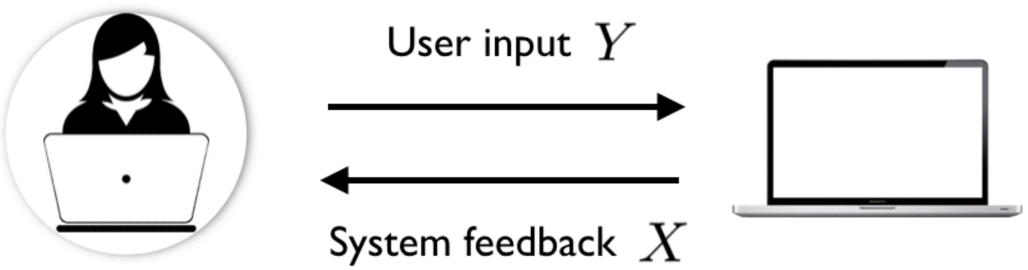


1950 **Experimental Psychology Applications**

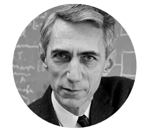
Hick's law (1952)

Fitts' law (1954)

2017 **Bayesian Information Gain**



1948 **Information Theory**

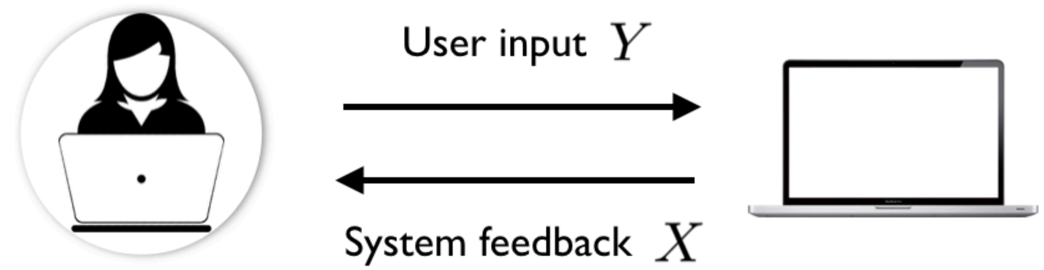


1950 **Experimental Psychology Applications**

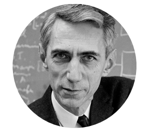
Hick's law (1952)

Fitts' law (1954)

2017 **Bayesian Information Gain**



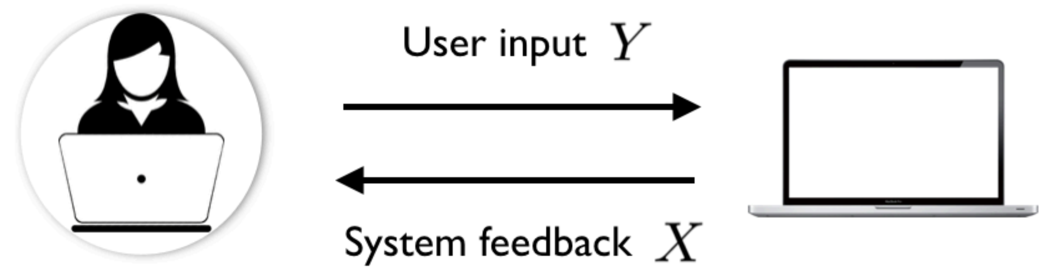
BIGnav (2017)

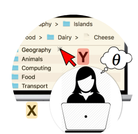
1948 **Information Theory** 

1950 **Experimental Psychology Applications**

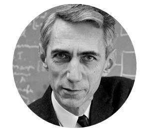
- **Hick's law** (1952) 
- **Fitts' law** (1954) 

2017 **Bayesian Information Gain**



- **BIGnav** (2017) 
- **BIGFile** (2018) 

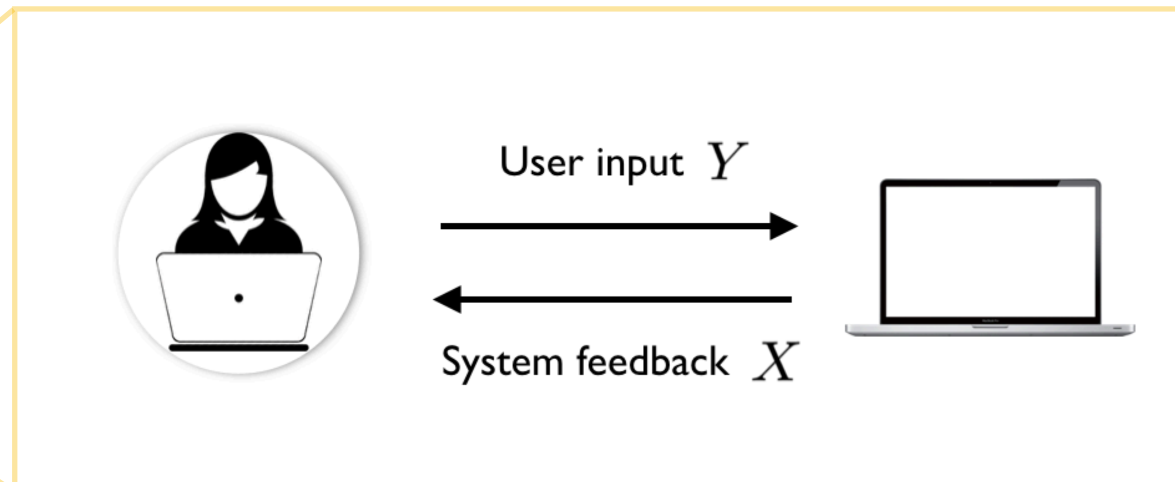
Overview

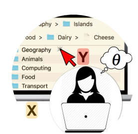
1948 **Information Theory** 

1950 **Experimental Psychology Applications**

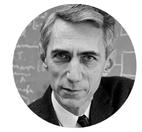
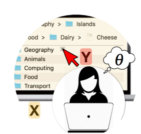
- **Hick's law** (1952) 
- **Fitts' law** (1954) 

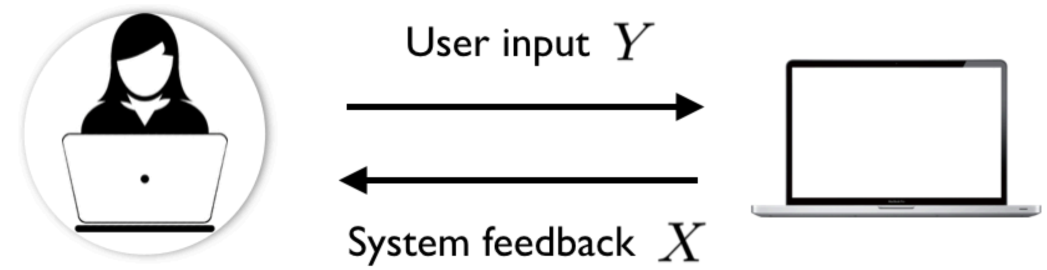
2017 **Bayesian Information Gain**

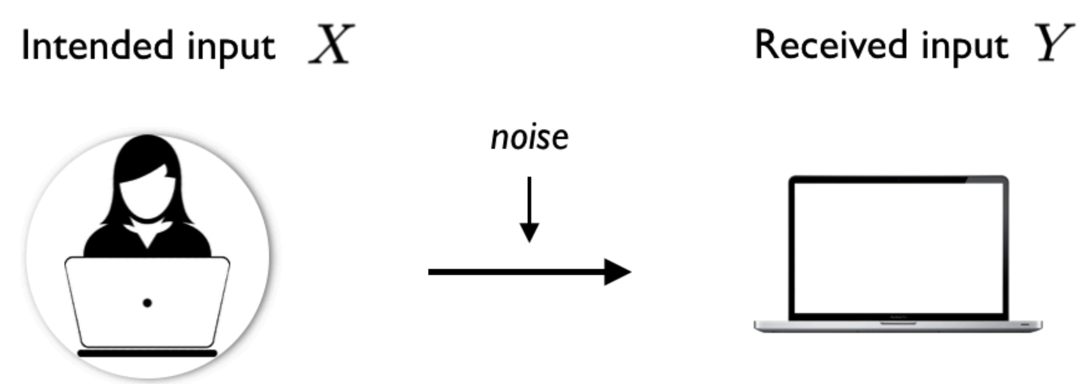


- **BIGnav** (2017) 
- **BIGFile** (2018) 

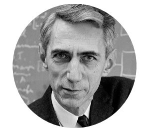
2018 **Information-Theoretic Measures**

- **1948** Information Theory 
- **1950** Experimental Psychology Applications
  - Hick's law (1952) 
  - Fitts' law (1954) 
- **2017** Bayesian Information Gain
  - BIGnav (2017) 
  - BIGFile (2018) 
- **2018** Information-Theoretic Measures





1948 **Information Theory**

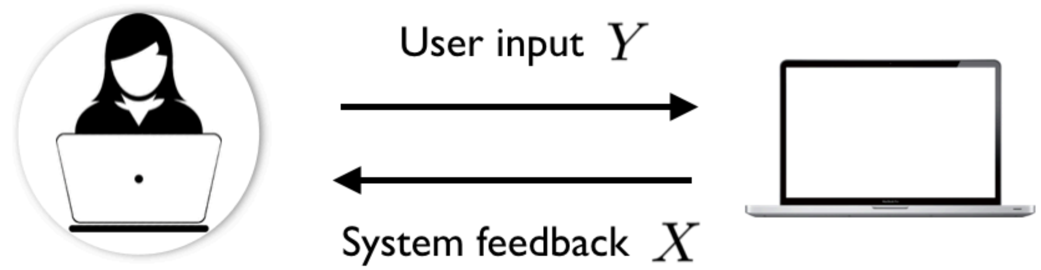


1950 **Experimental Psychology Applications**

Hick's law (1952)

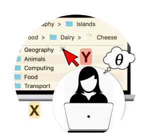
Fitts' law (1954)

2017 **Bayesian Information Gain**

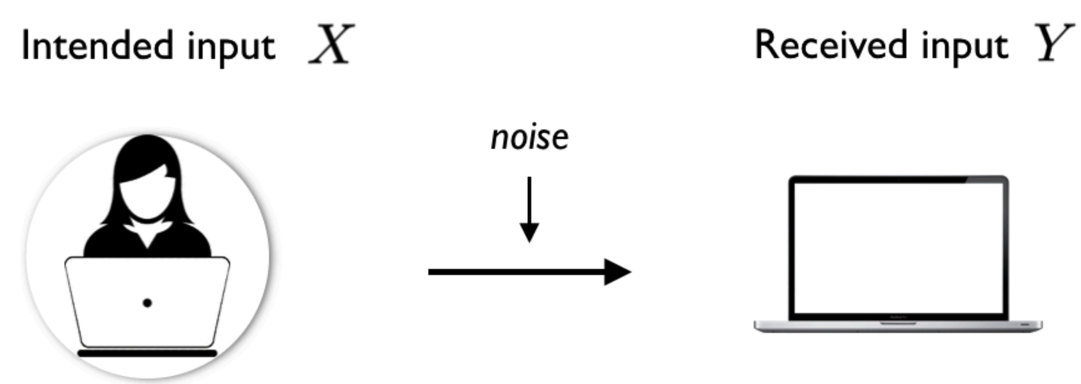


BIGnav (2017)

BIGFile (2018)

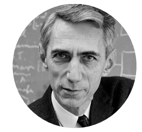


2018 **Information-Theoretic Measures**



Command Selection

1948 **Information Theory**

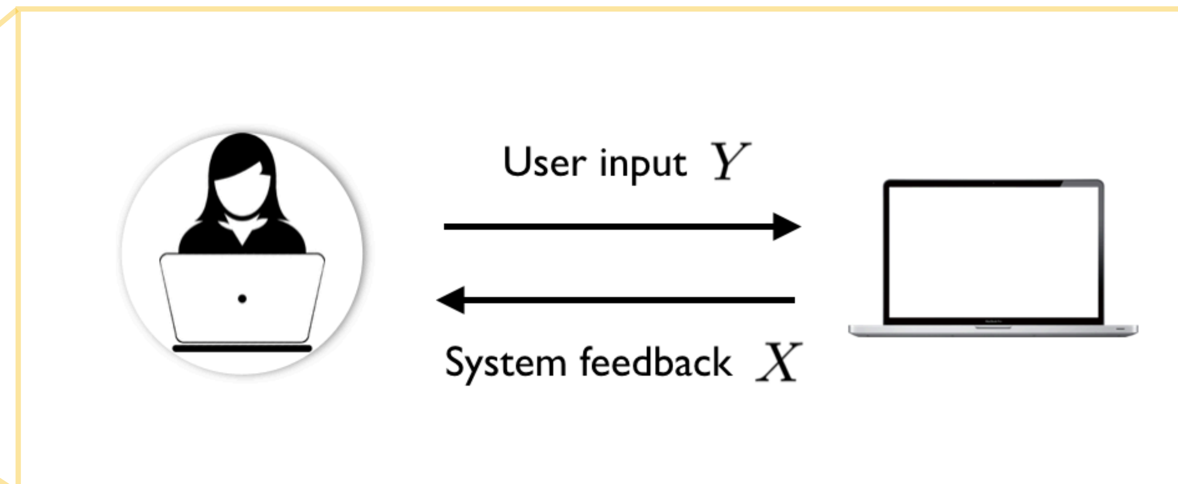


1950 **Experimental Psychology Applications**

Hick's law (1952)

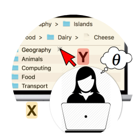
Fitts' law (1954)

2017 **Bayesian Information Gain**

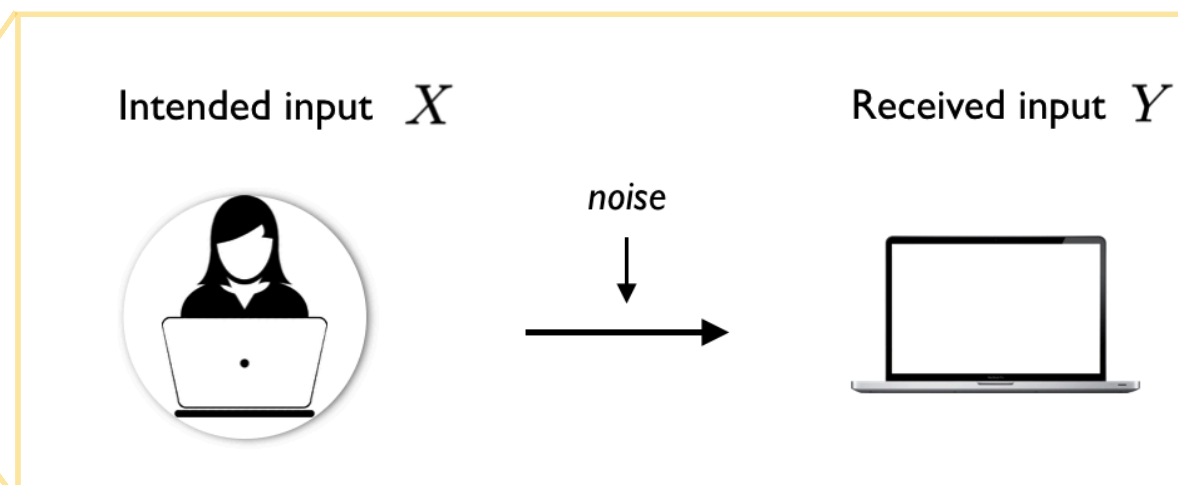


BIGnav (2017)

BIGFile (2018)

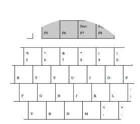


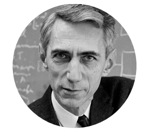
2018 **Information-Theoretic Measures**



Command Selection

Text Entry

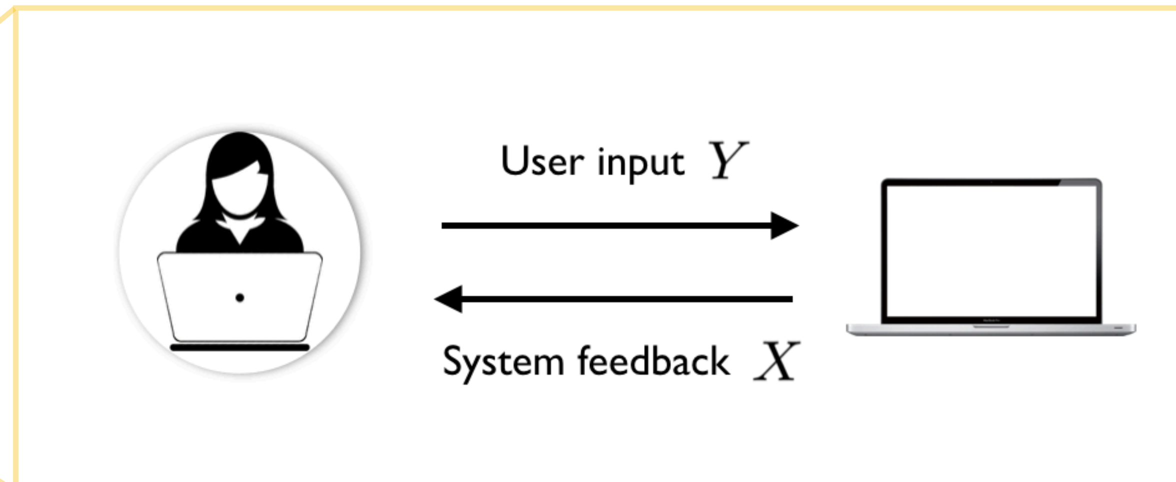


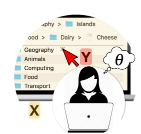
1948 **Information Theory** 

1950 **Experimental Psychology Applications**

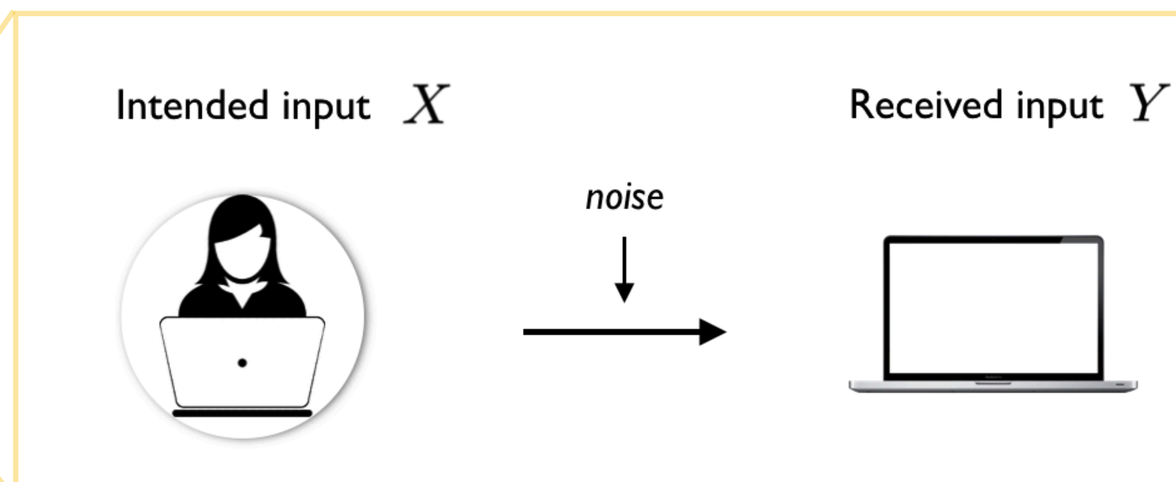
- **Hick's law** (1952) 
- **Fitts' law** (1954) 

2017 **Bayesian Information Gain**

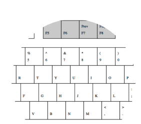


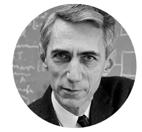
- **BIGnav** (2017) 
- **BIGFile** (2018) 

2018 **Information-Theoretic Measures**



- **Command Selection** 

- **Text Entry** 

1948 **Information Theory** 

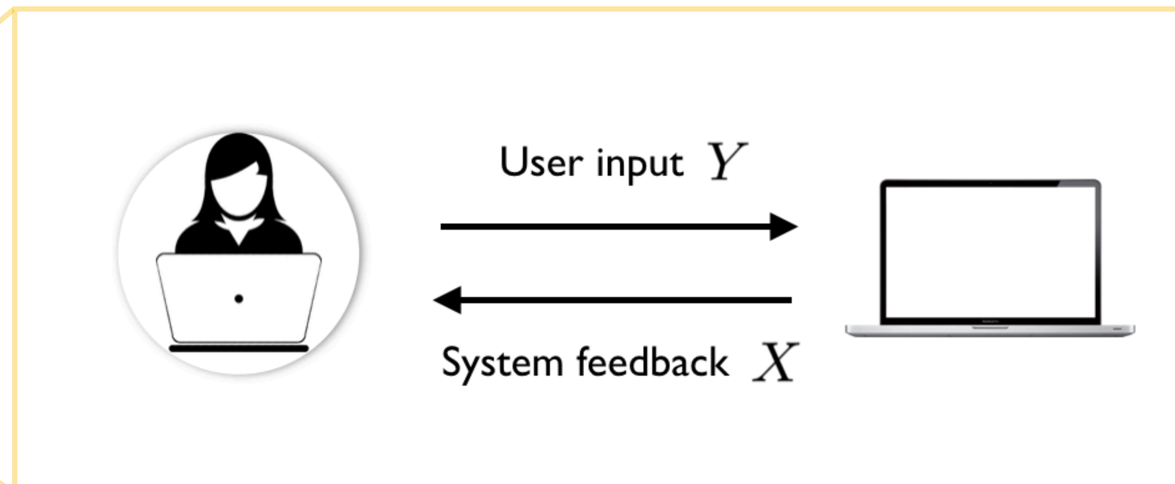
Part i

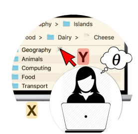
1950 **Experimental Psychology Applications**

- **Hick's law** (1952) 
- **Fitts' law** (1954) 

Part ii

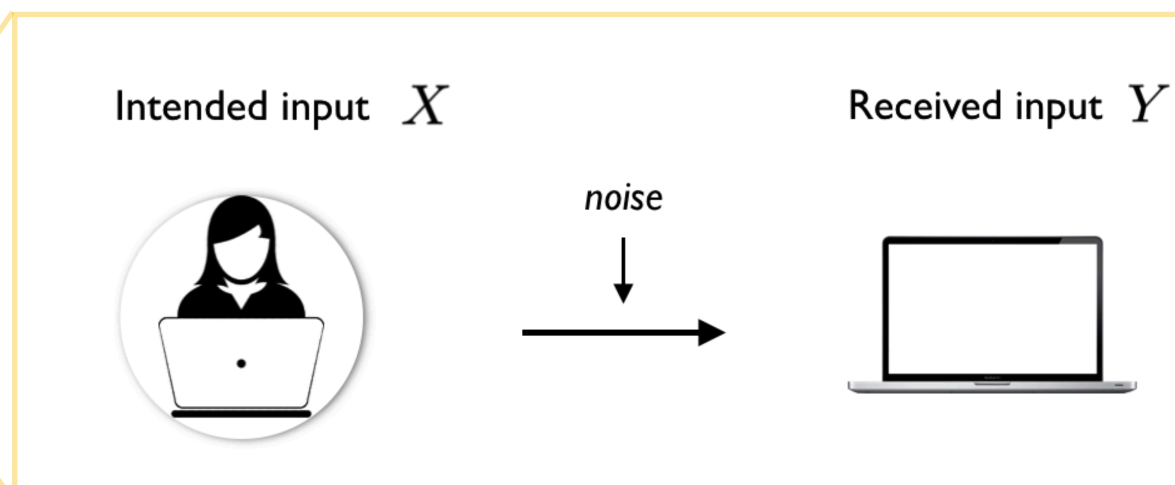
2017 **Bayesian Information Gain**



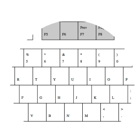
- **BIGnav** (2017) 
- **BIGFile** (2018) 

Part iii

2018 **Information-Theoretic Measures**



- **Command Selection** 

- **Text Entry** 

1948 Information Theory

1950 Experimental Psychology Applications

Hick's law (1952)

Part i

Fitts' law (1954)

Whatever Happened to Information Theory in Psychology? (Luce 2003)

2017 Bayesian Information Gain

BIGnav (2017)

Part ii

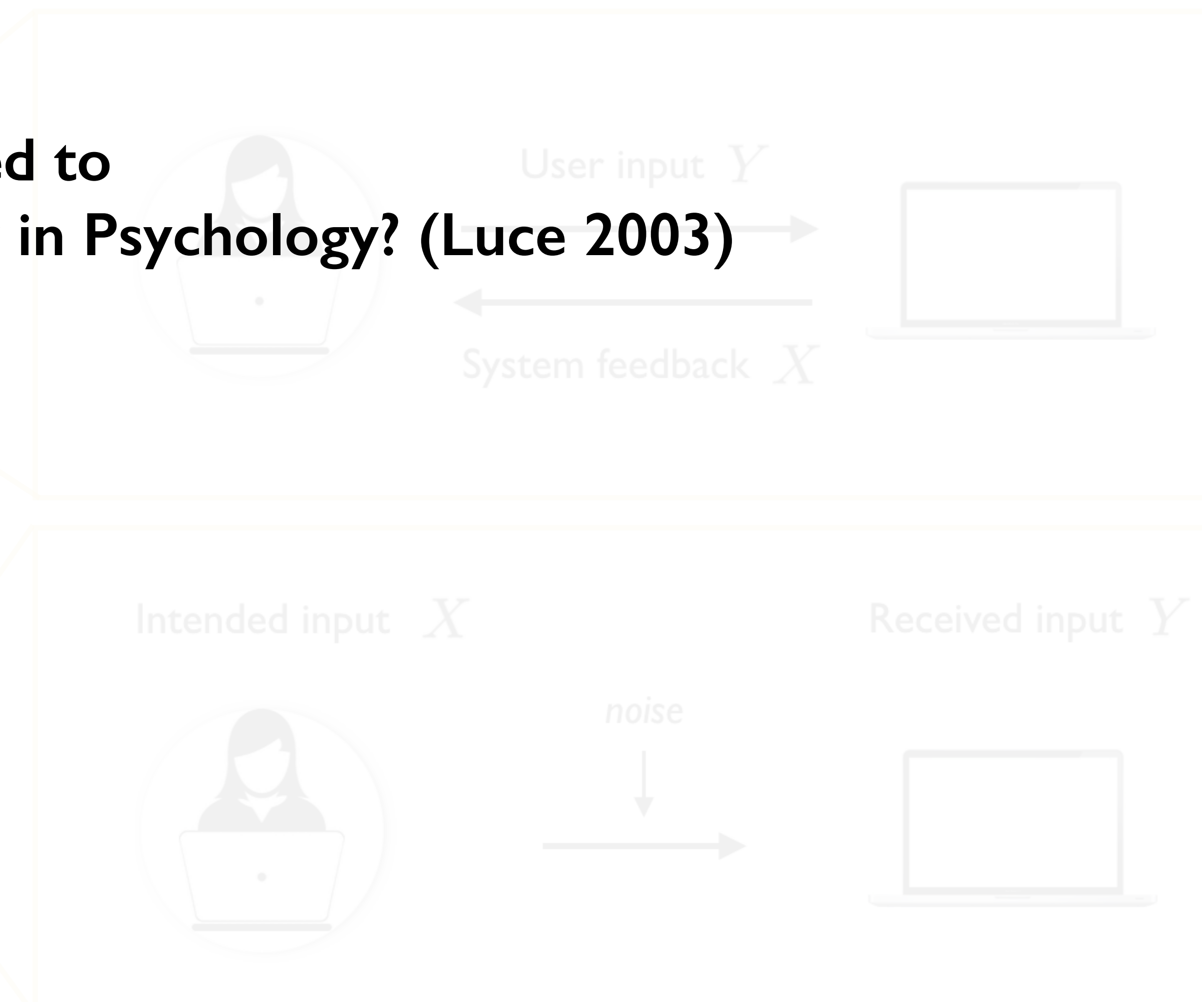
BIGFile (2018)

2018 Information-Theoretic Measures

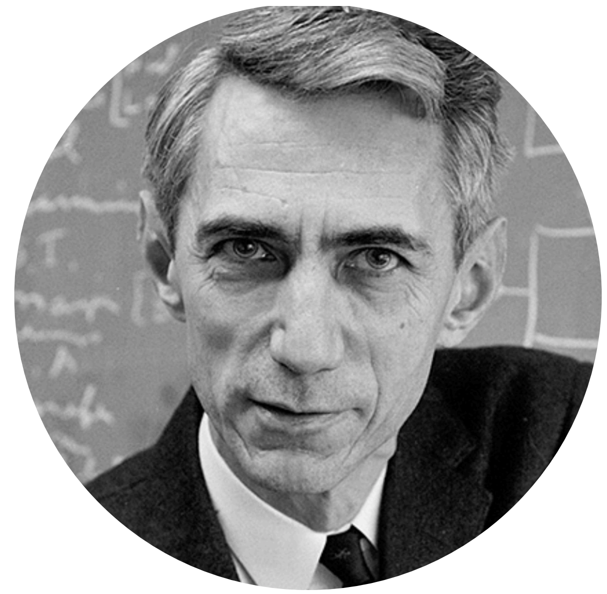
Command Selection

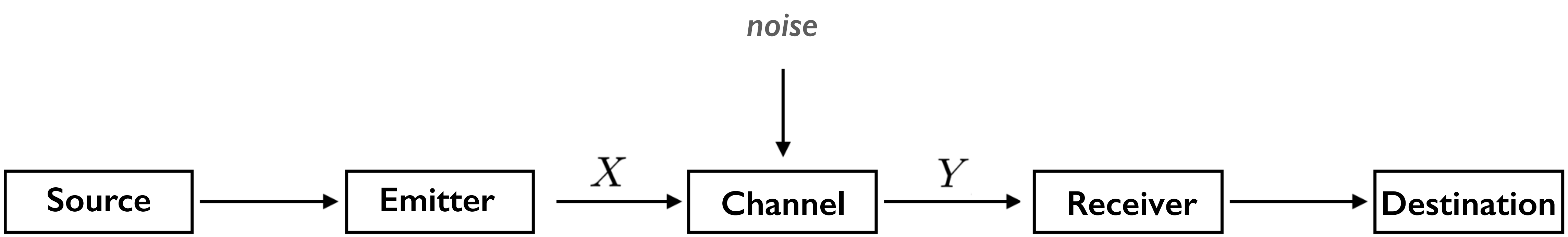
Part iii

Text Entry



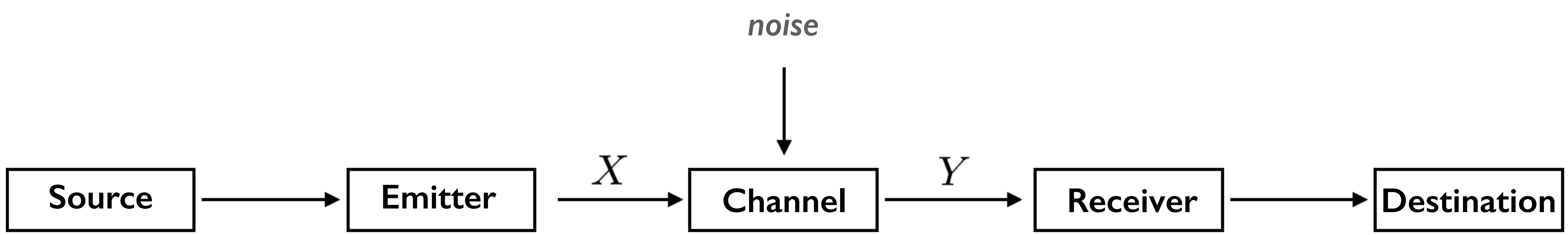
Background





A mathematical theory of communication (Shannon 1948)

Background

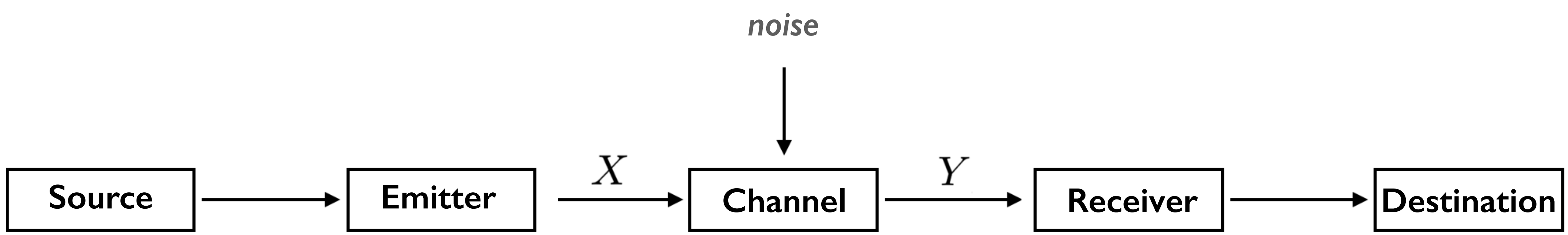


Discrete random variable X:

$$\{x_1, x_2, \ldots, x_n\}$$

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

Background

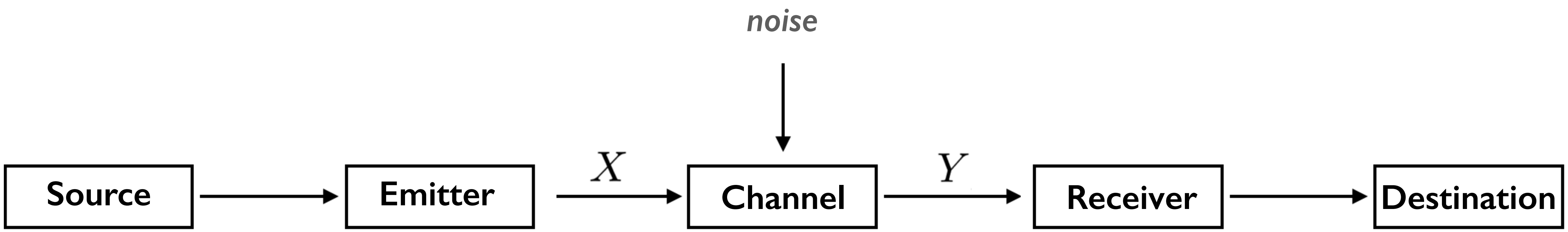


Discrete random variable X:

$$\{x_1, x_2, \ldots, x_n\}$$

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

Background



Discrete random variable X:

$$\{x_1, x_2, \ldots, x_n\}$$

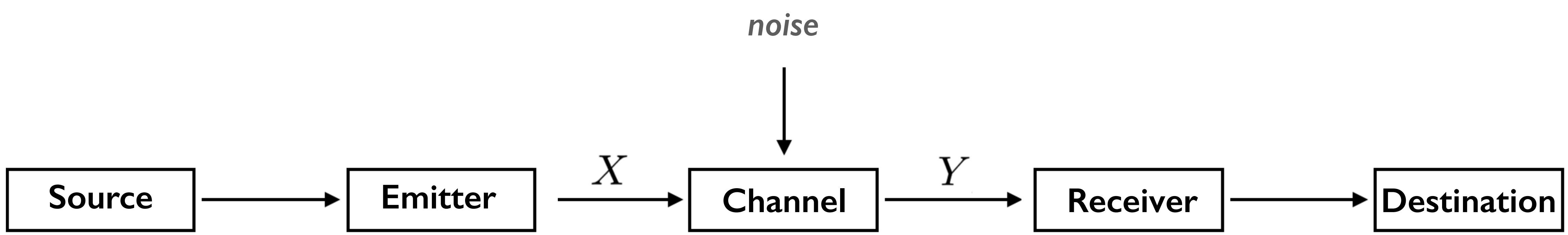
$$P_1 \quad P_2 \quad \ldots \quad P_n$$

Information as entropy:

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

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

Background



Discrete random variable X:

$$\{x_1, x_2, \ldots, x_n\}$$

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

Information as entropy:

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

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

Discrete random variable X:

$$\{x_1, x_2, \ldots, x_n\}$$

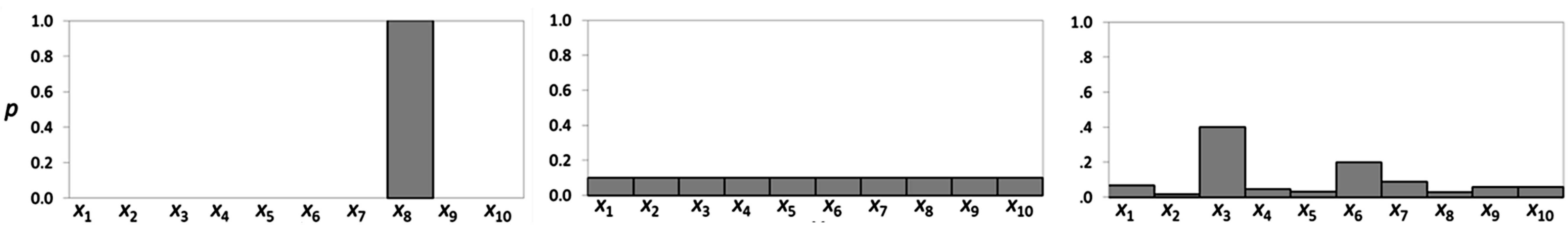
$$P_1 \quad P_2 \quad \ldots \quad P_n$$

Information as entropy:

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

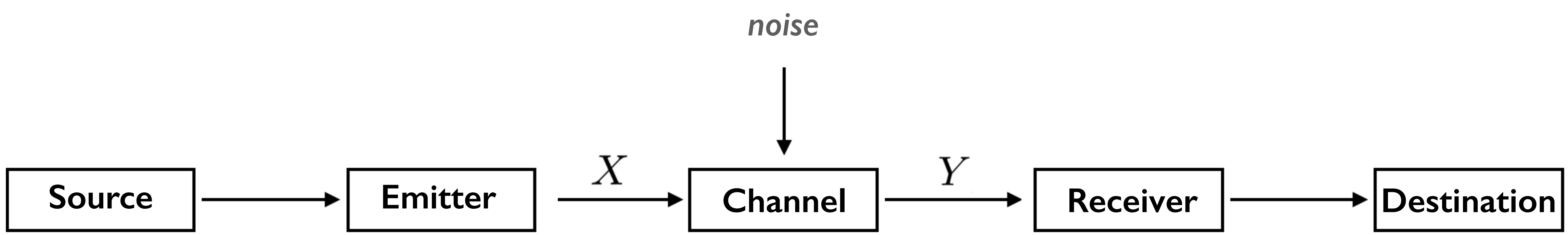
$$0 \le H(X) \le \log N$$

$H(X) = 0$ **bits** $\qquad H(X) = \log_2 10 = 3.3$ **bits** $\qquad 0 < H(X) = 2.7 < 3.3$ **bits**



Not Just Pointing: Shannon's Information Theory as a General Tool for Performance Evaluation of Input Techniques (Guiard 2018)

Background



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

- Choice-reaction time (Hick, Hyman)
- Information Capacity of Motor Movement (Fitts)
- Information Capacity of Working Memory (Miller)

The Magical Number Seven, Plus or Minus Two (Miller 1956)

Choice-reaction time

Helmholtz
(1821-1894)

Choice-reaction time

Donders 1868:
First report of choice-reaction times

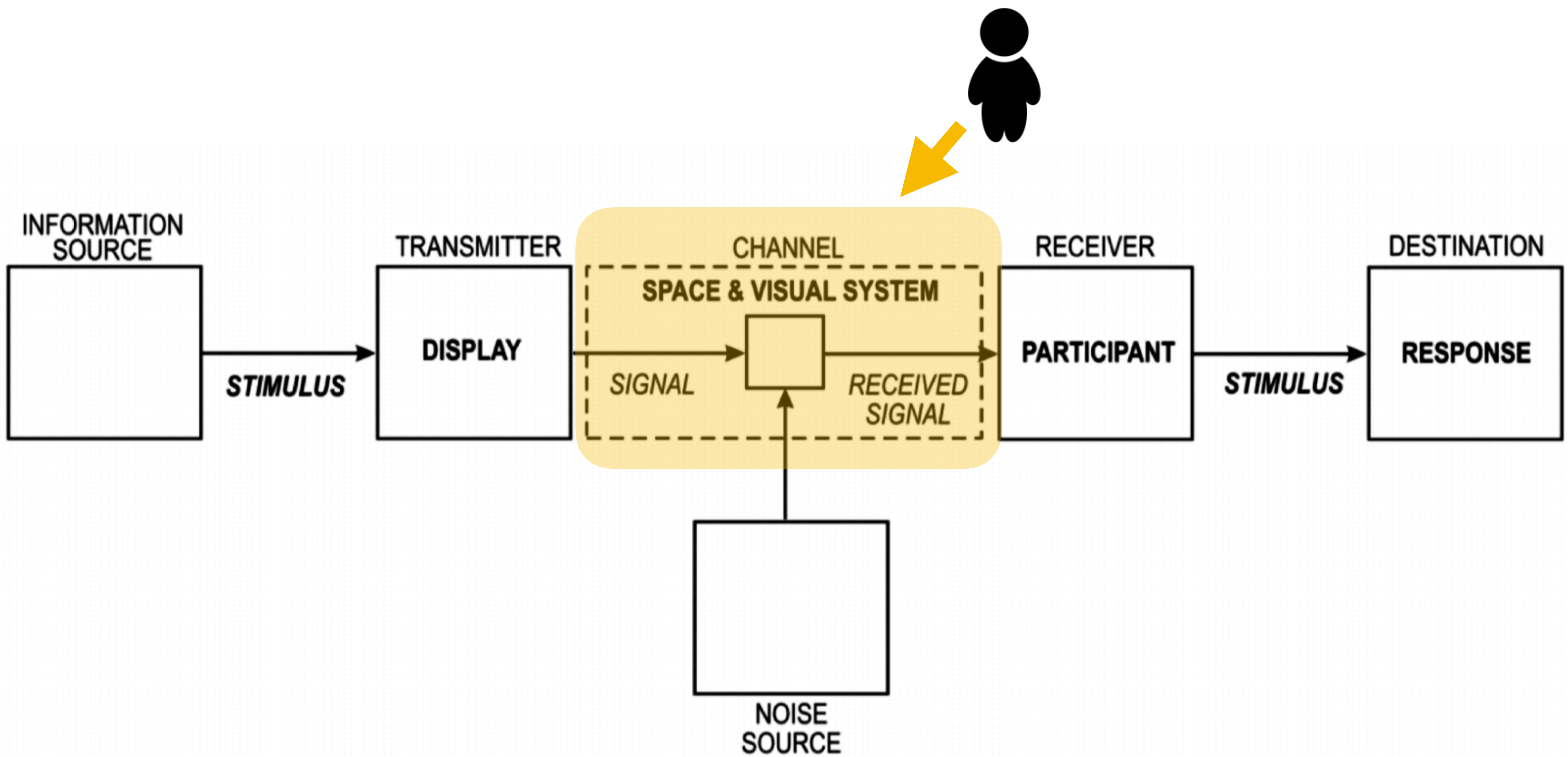
Merkel 1883:
Relationship between time and the number of stimuli (linear)

Blank 1934:
A logarithmic relationship is mentioned

Shannon 1948:
Information Theory

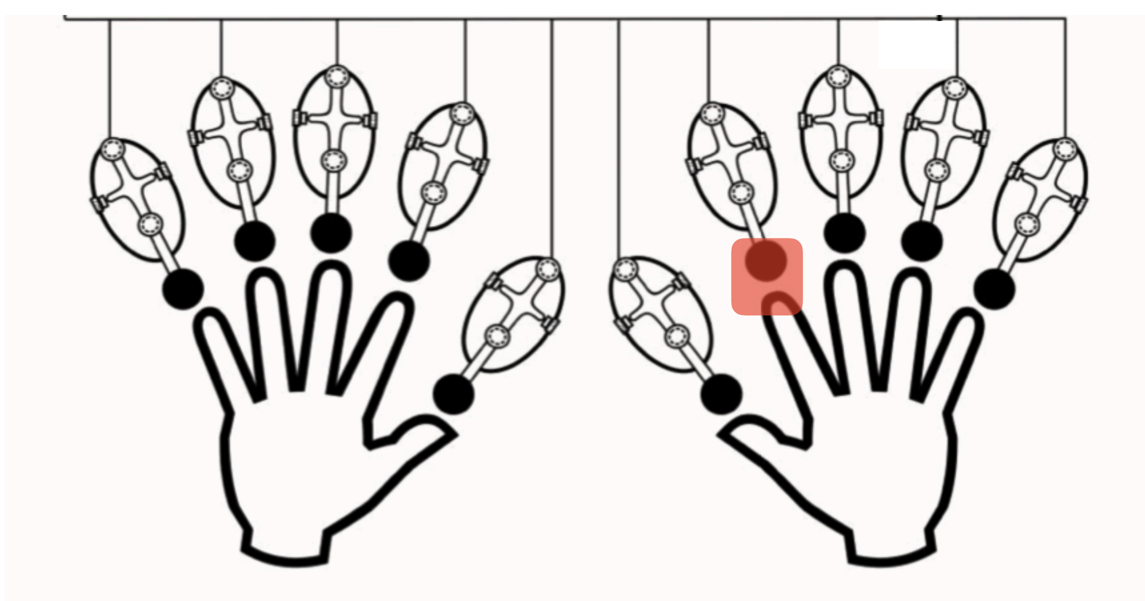
Hick 1952
Hyman 1953

Choice-reaction time (Hick, Hyman)

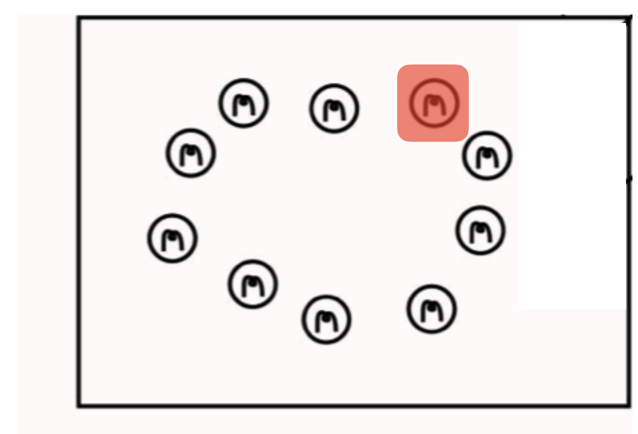


Information theory of choice-reaction times (Laming 1968)

Choice-reaction time (Hick, Hyman)



Morse keys

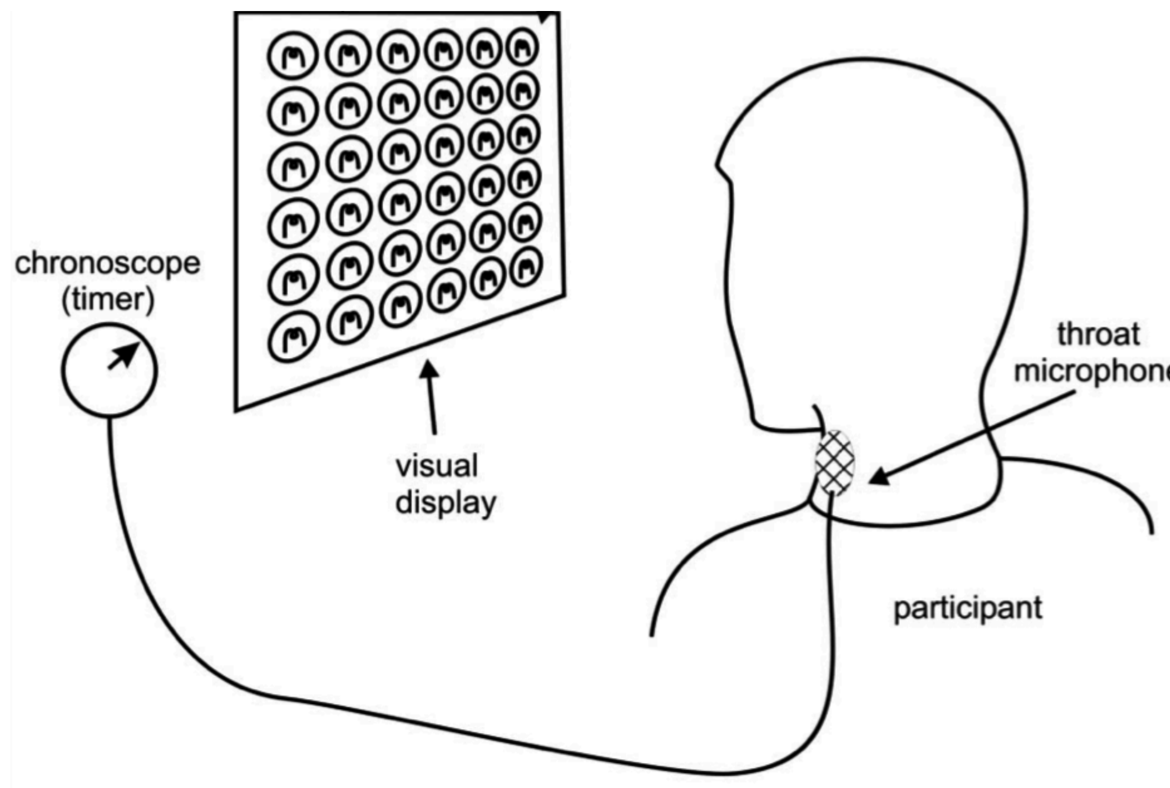


Pea lamps

$$T \propto \log_2 n$$

On the rate of gain of gain of information (Hick 1952)

Choice-reaction time (Hick, Hyman)



$$T \propto \sum_{i=1}^{n} p_i \log_2(1/p_i)$$

Stimulus information as a determinant of reaction time (Hyman 1953)

Choice-reaction time (Hick, Hyman, etc.)

	Hick	*Hyman*	*Landauer & Nachbar*	*Cockburn et al.*	*Soukoreff & Mackenzie*	*Mackenzie et al.*	*Wobbrock & Myers*
Task	Reaction	Reaction	VS	Decision	VS	VS	VS
Stimuli	Random	Random	Ordered	Random	Keyboard	Keyboard	Random
Participants	Well-trained	Well-trained	All users	Users starting from block 2	Novice users	Novice users	All users
Distribution	Uniform	Non-uniform	Uniform	Zipfian	Uniform	Uniform	Uniform
Information	Up to 3.32 bits	Up to 2.81 bits	Up to 4 bits	Up to 3.58 bits	4.75 bits	4.75 bits	Up to 2 bits
Measure	Mutual Information	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy
Formula	$\log(n+1)$ or $\log(n_e+1)$	$\log n$ or $-\sum_{i=1}^{n} p_i \log_2 p_i$	$\log n$	$\log n$	$\log n$	$\log n$	$\log n$
Results	Logarithmic	Logarithmic	Logarithmic	Logarithmic	Logarithmic	Not Logarithmic	Logarithmic

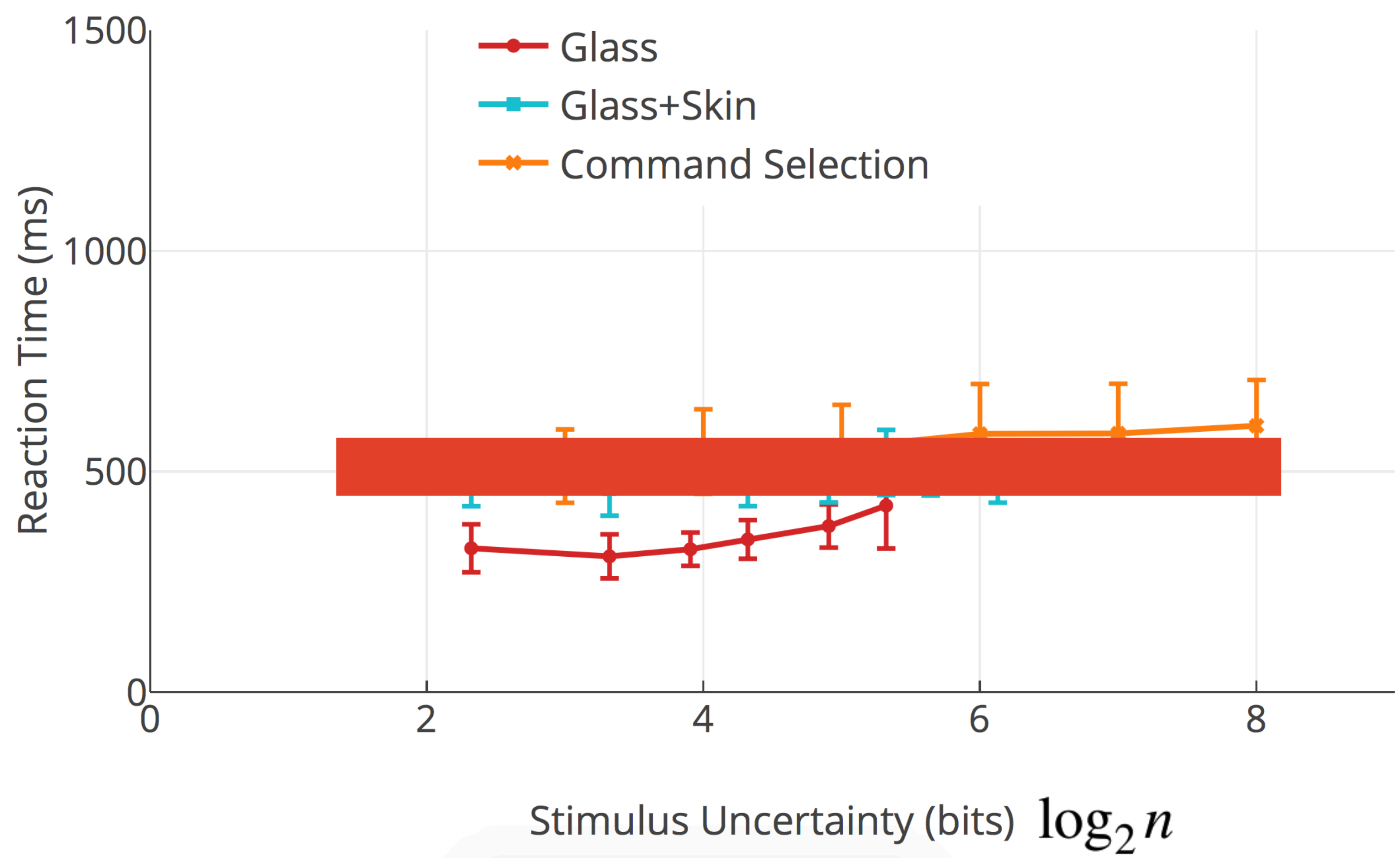
- VS: Visual search.

Choice-reaction time (Hick, Hyman, etc.)

	Hick	*Hyman*	*Landauer & Nachbar*	*Cockburn et al.*	*Soukoreff & Mackenzie*	*Mackenzie et al.*	*Wobbrock & Myers*
Task	Reaction	Reaction	VS	Decision	VS	VS	VS
Stimuli	Random	Random	Ordered	Random	Keyboard	Keyboard	Random
Participants	Well-trained	Well-trained	All users	Users starting from block 2	Novice users	Novice users	All users
Distribution	Uniform	Non-uniform	Uniform	Zipfian	Uniform	Uniform	Uniform
Information	Up to 3.32 bits	Up to 2.81 bits	Up to 4 bits	Up to 3.58 bits	4.75 bits	4.75 bits	Up to 2 bits
Measure	Mutual Information	Entropy	Entropy	Entropy	Entropy	Entropy	Entropy
Formula	$\log(n+1)$ or $\log(n_e+1)$	$\log n$ or $-\sum_{i=1}^{n} p_i \log_2 p_i$	$\log n$	$\log n$	$\log n$	$\log n$	$\log n$
Results	Logarithmic	Logarithmic	Logarithmic	Logarithmic	Logarithmic	Not Logarithmic	Logarithmic

- VS: Visual search.

Implications for HCI: Effect size of Hick's law is insignificant

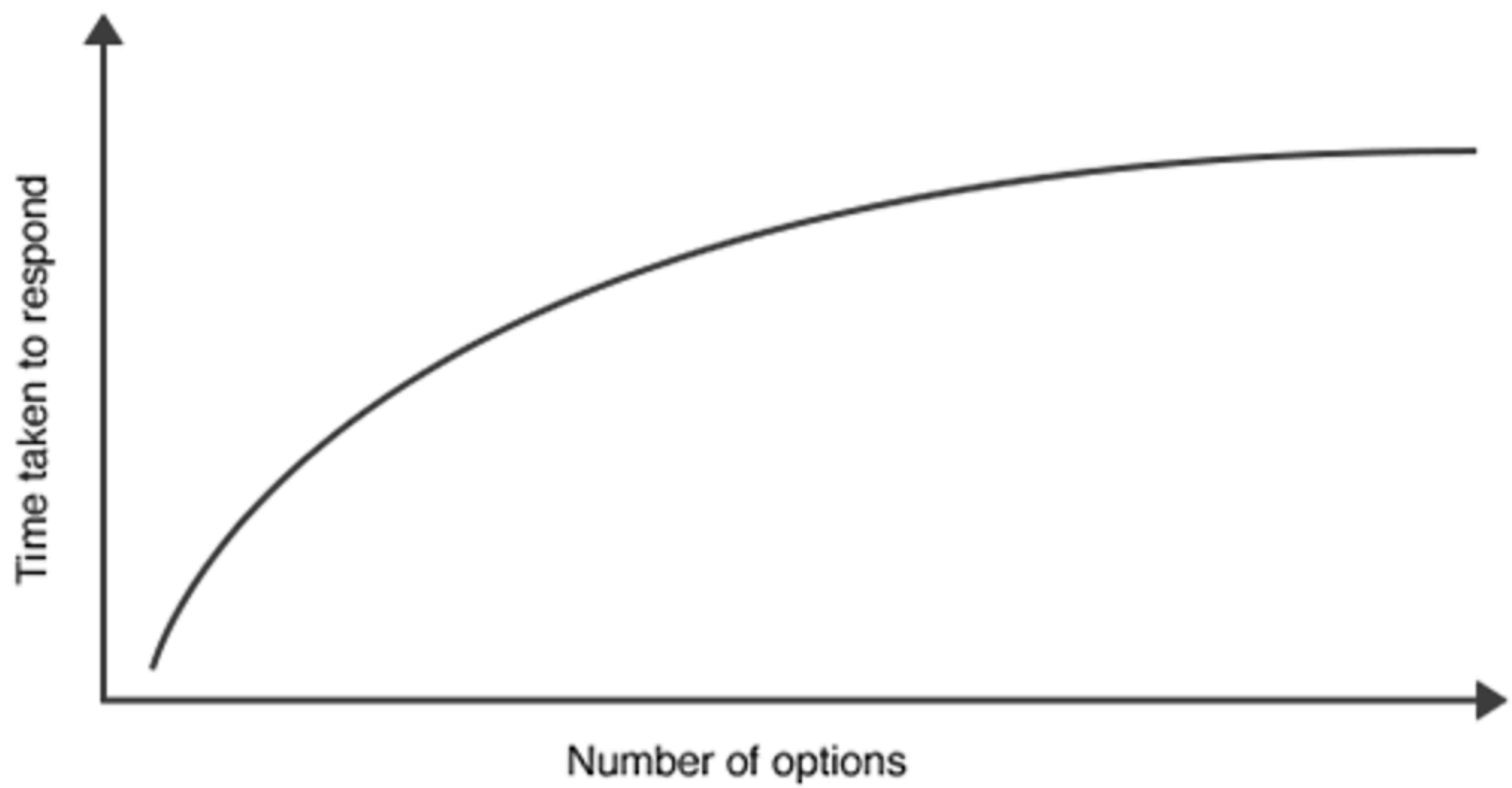


Wanyu Liu, Julien Gori, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard. How Relevant is Hick's Law for HCI? (CHI '19) [under review]

- Hick's law in design

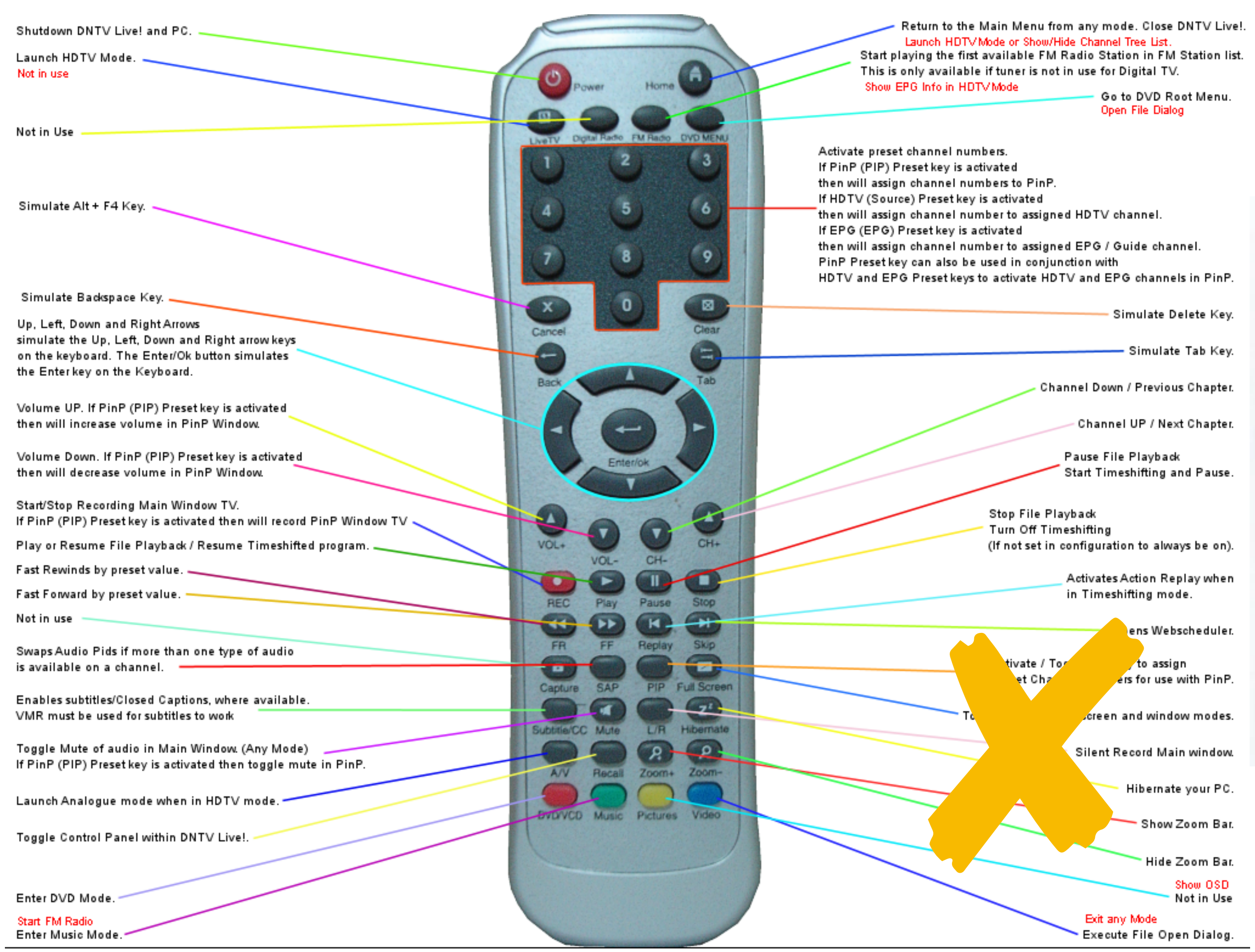
- Hick's Law: Making the choice easier for users
- Juggling Jam: Applying Hick's Law to Web Design
- Design principle: Hick's Law — quick decision making
- Universal Principles of Design (Lidwell 2010)

Hick's law in design



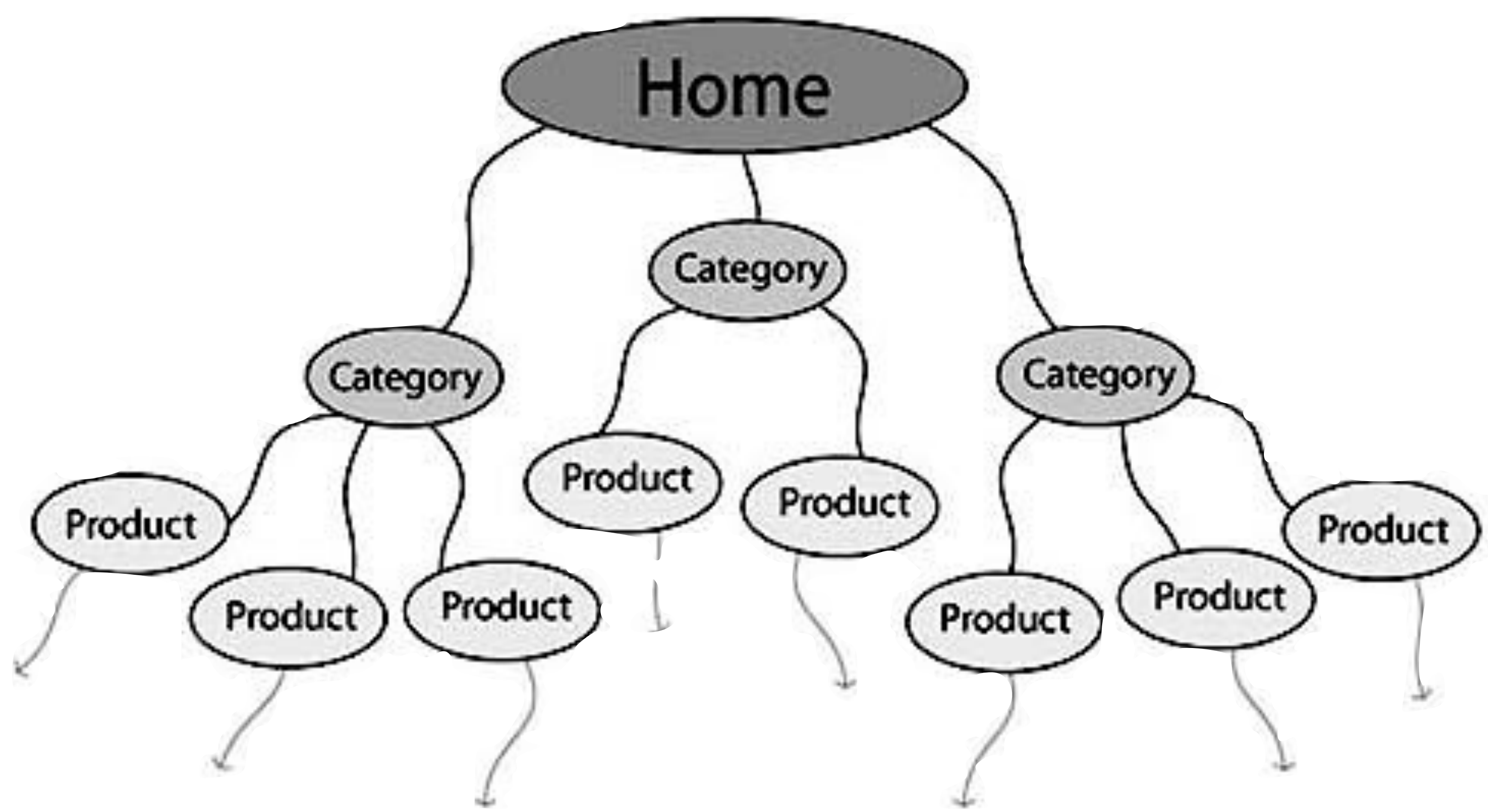
Hick’s law in design

Do not bombard users with choices



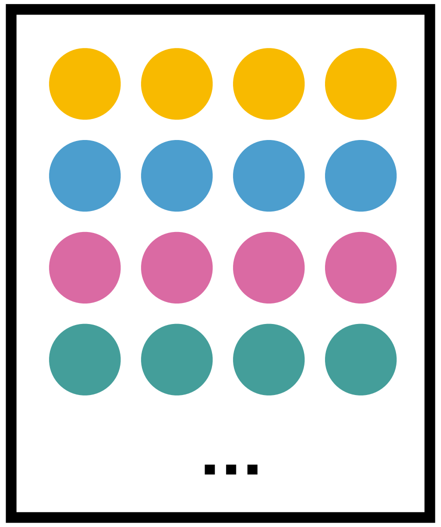
Hick's law in design

Always categorize choices



✻ Implications for HCI: Hick's law does not justify design "rules"

N = 512

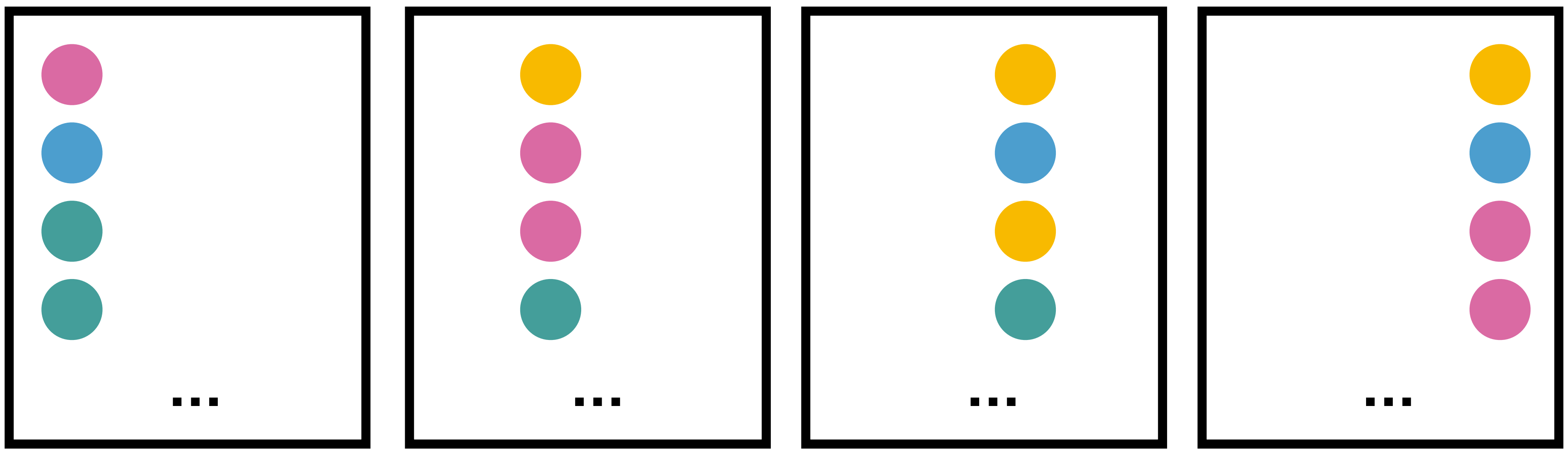


$$\mathrm{RT} = a + b\log_2(512) = a + 9 \times b$$

Wanyu Liu, Julien Gori, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard.
How Relevant is Hick's Law for HCI? (CHI '19) [under review]

Implications for HCI: Hick's law does not justify design "rules"

N = 512

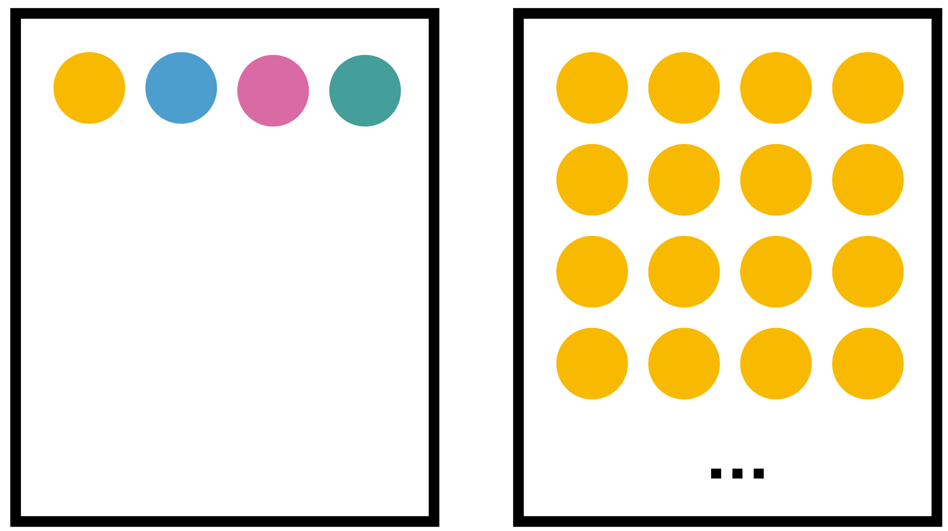


$$\mathrm{RT} = 1/4\Big(\sum_{i=1}^{4} a\ i + b\ i\ \log_2(128)\Big) = 10/4\ a + 70/4\ b$$

Wanyu Liu, Julien Gori, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard. How Relevant is Hick's Law for HCI? (CHI '19) [under review]

Implications for HCI: Hick's law does not justify design "rules"

N = 512

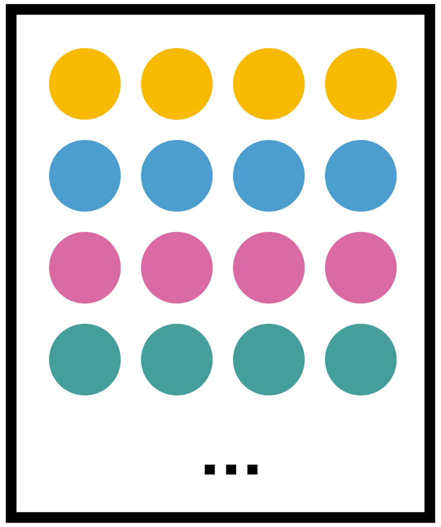


$$\mathrm{RT} = a + b\log_2(4) + a + b\log_2(128) = 2a + 9 \times b$$

Wanyu Liu, Julien Gori, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard.
How Relevant is Hick's Law for HCI? (CHI '19) [under review]

Implications for HCI: Hick's law does not justify design "rules"

N = 512



$$\mathrm{RT} = a + b\log_2(512) = a + 9 \times b$$

Wanyu Liu, Julien Gori, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard. How Relevant is Hick's Law for HCI? (CHI '19) [under review]

Because human-computer interaction studies a human and a machine in **communication**, it draws from supporting knowledge on both the machine and the human side.

User

communication

Computer

ACM SIGCHI Curriculum for human-computer interaction (2009)

User

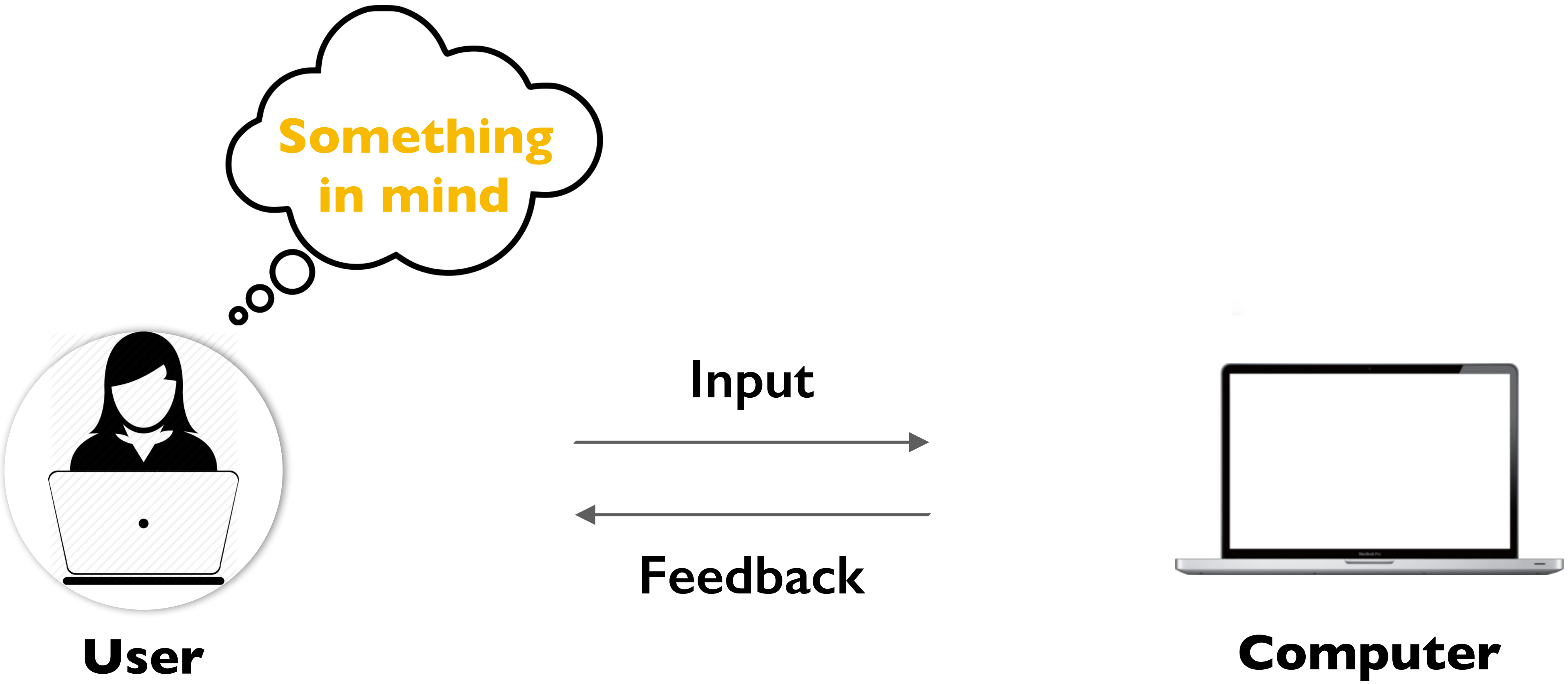
communication

Computer

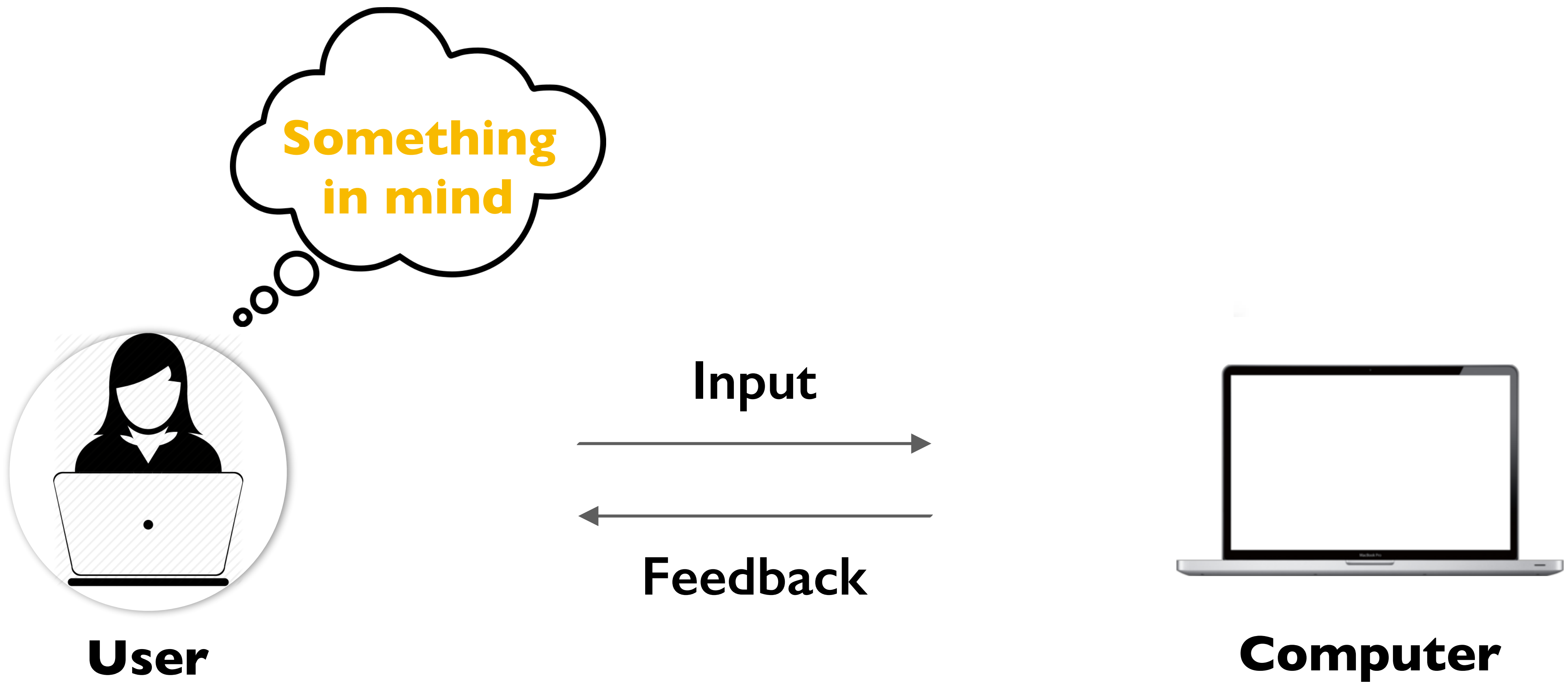
Part ii: Bayesian Information Gain (BIG)

User

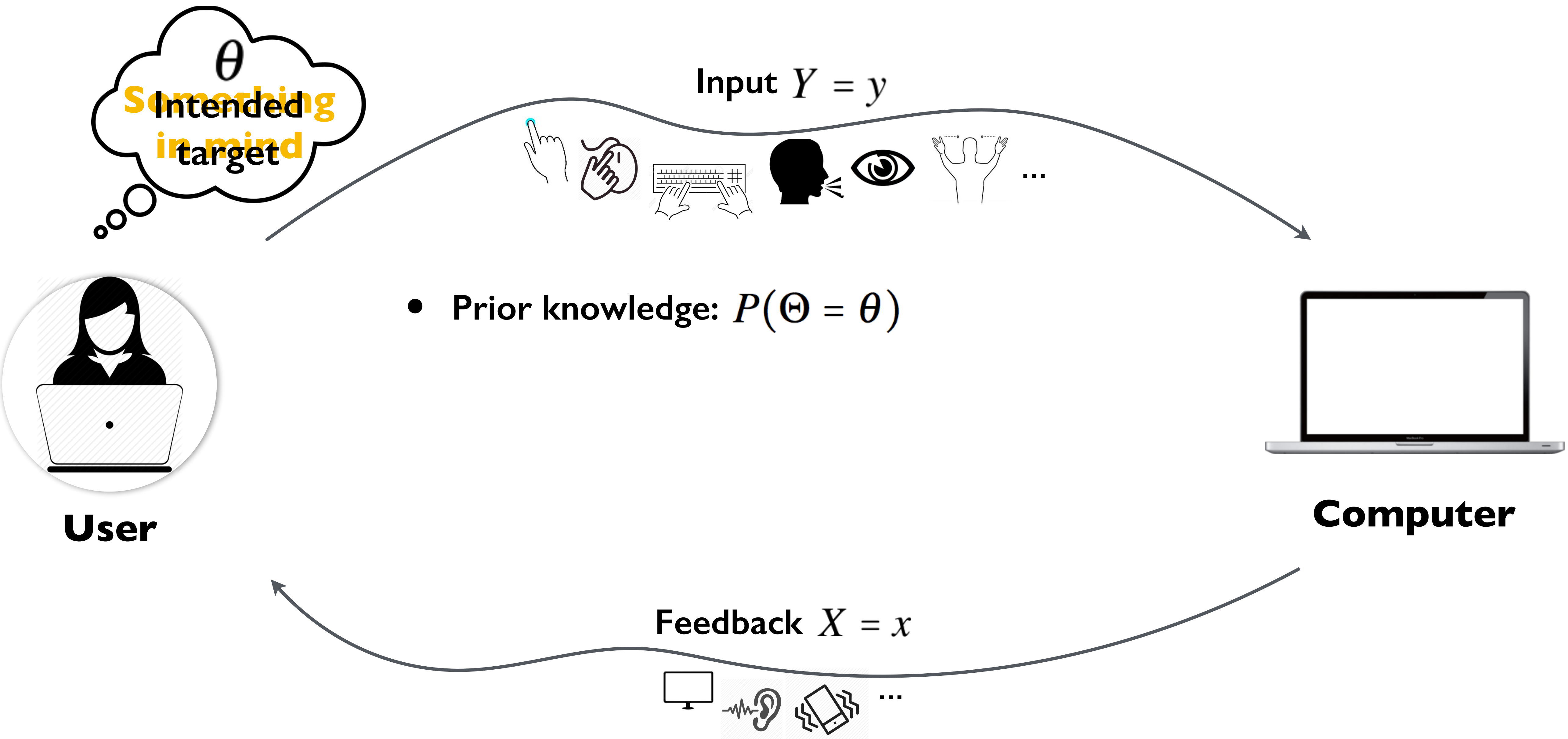
Computer



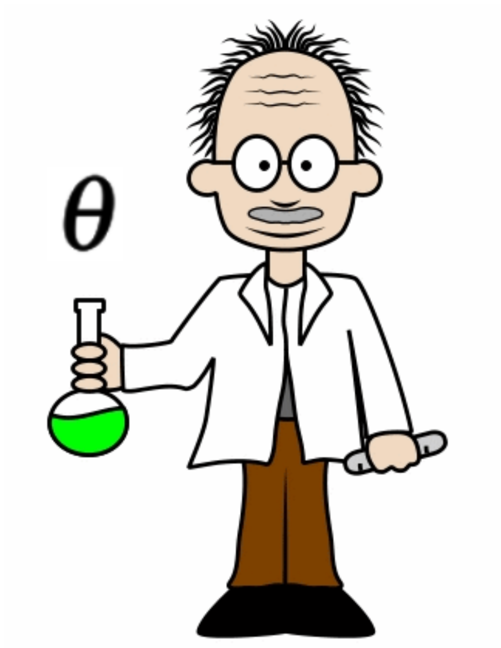
- To look for a restaurant
- To type a word
- To draw a gesture
- To select an icon
- To do something
- ....



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



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



θ

The Scientist

Experiment

x_1

x_2

x_3

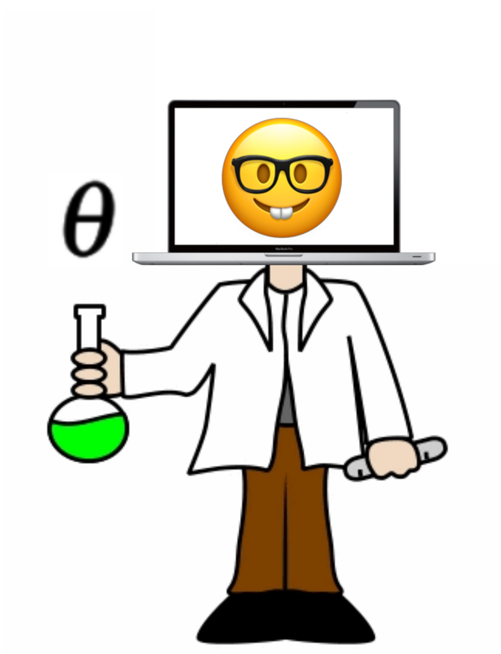
Observation

y_1

y_2

y_3

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



θ

The computer

Experiment

x_1

x_2

x_3

Observation

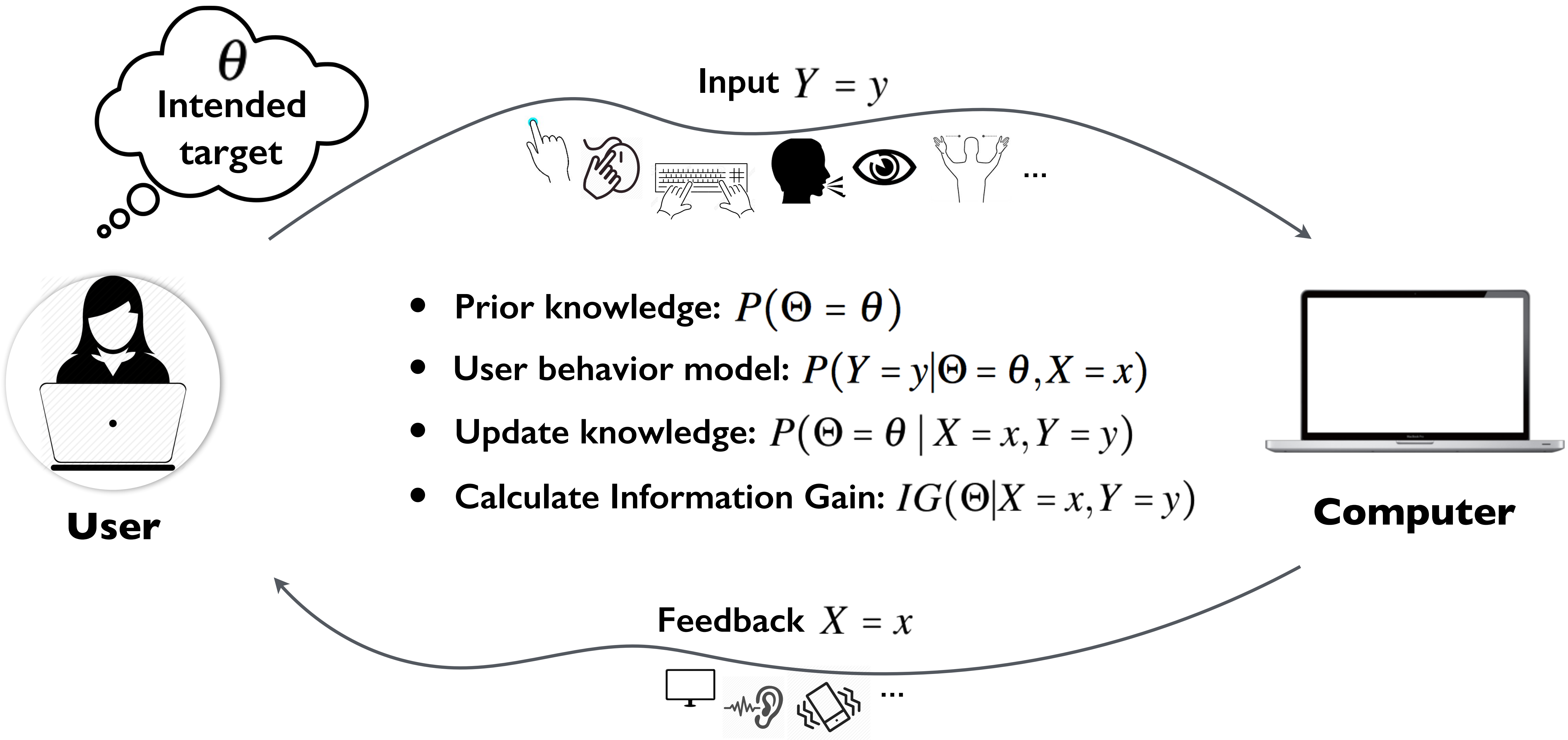
y_1

y_2

y_3

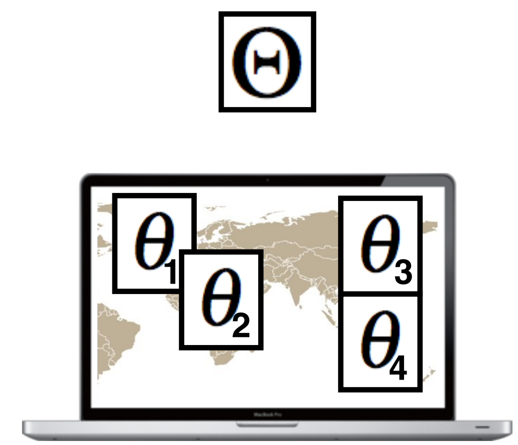
On a Measure of the Information Provided by an Experiment (Lindley 1956)

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



Bayesian Information Gain

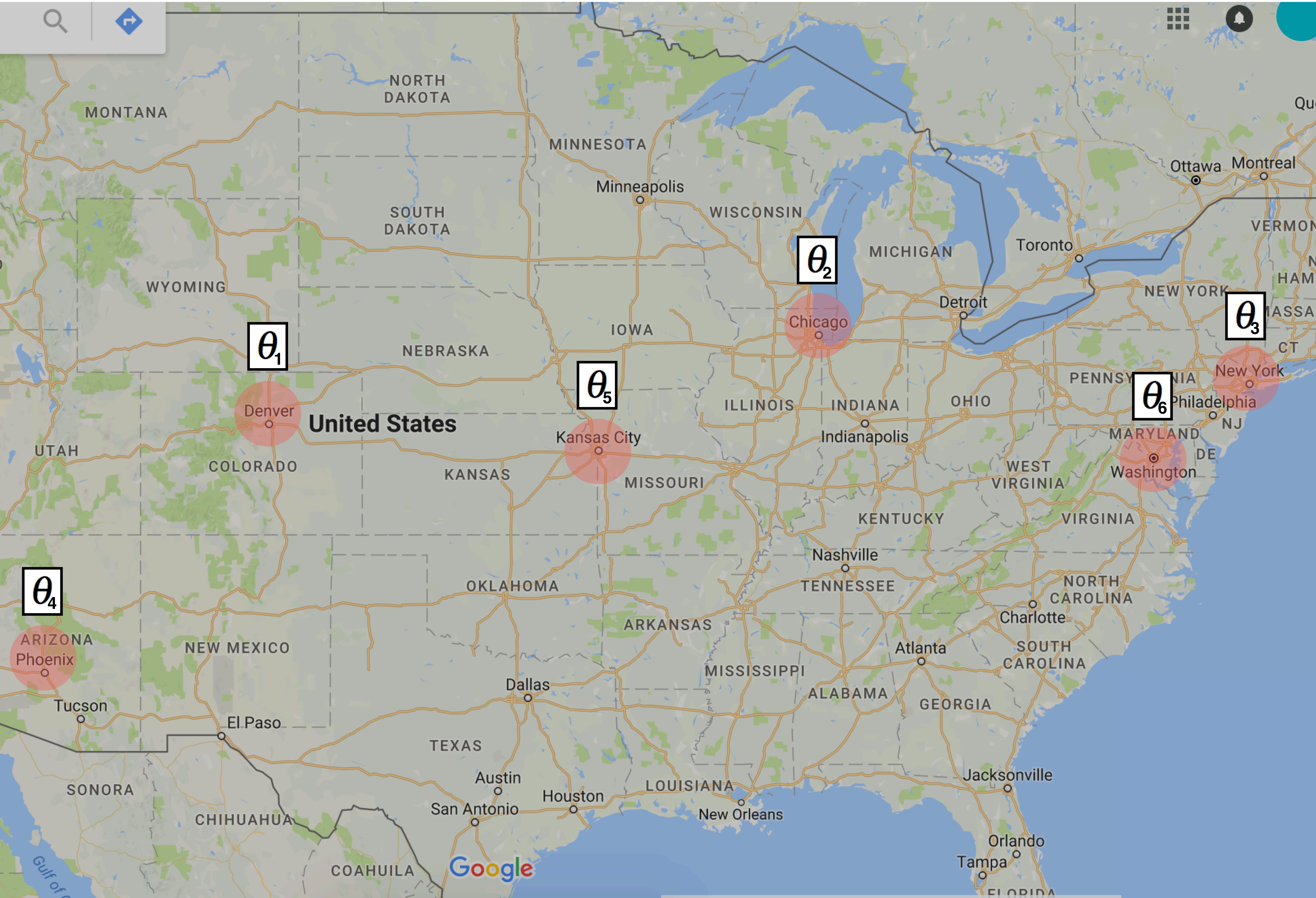
BIGnav



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

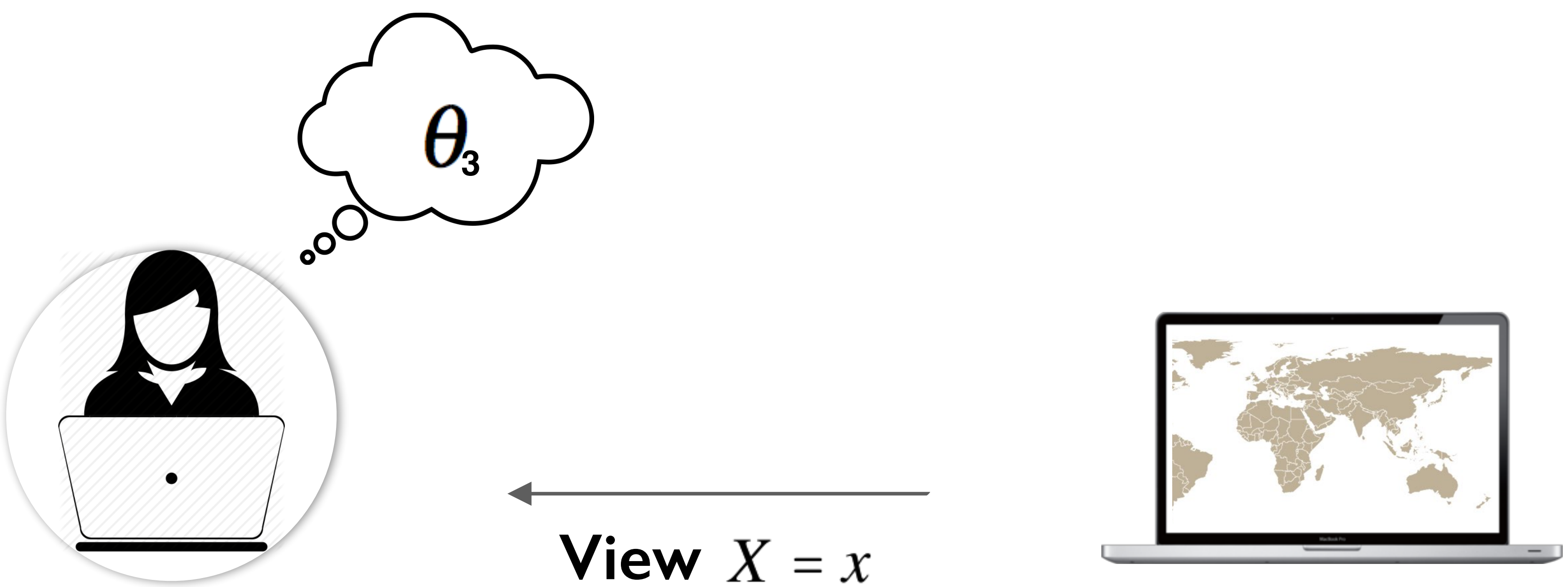
1948 — 1950 — 2017 **Bayesian Information Gain**

BIGnav



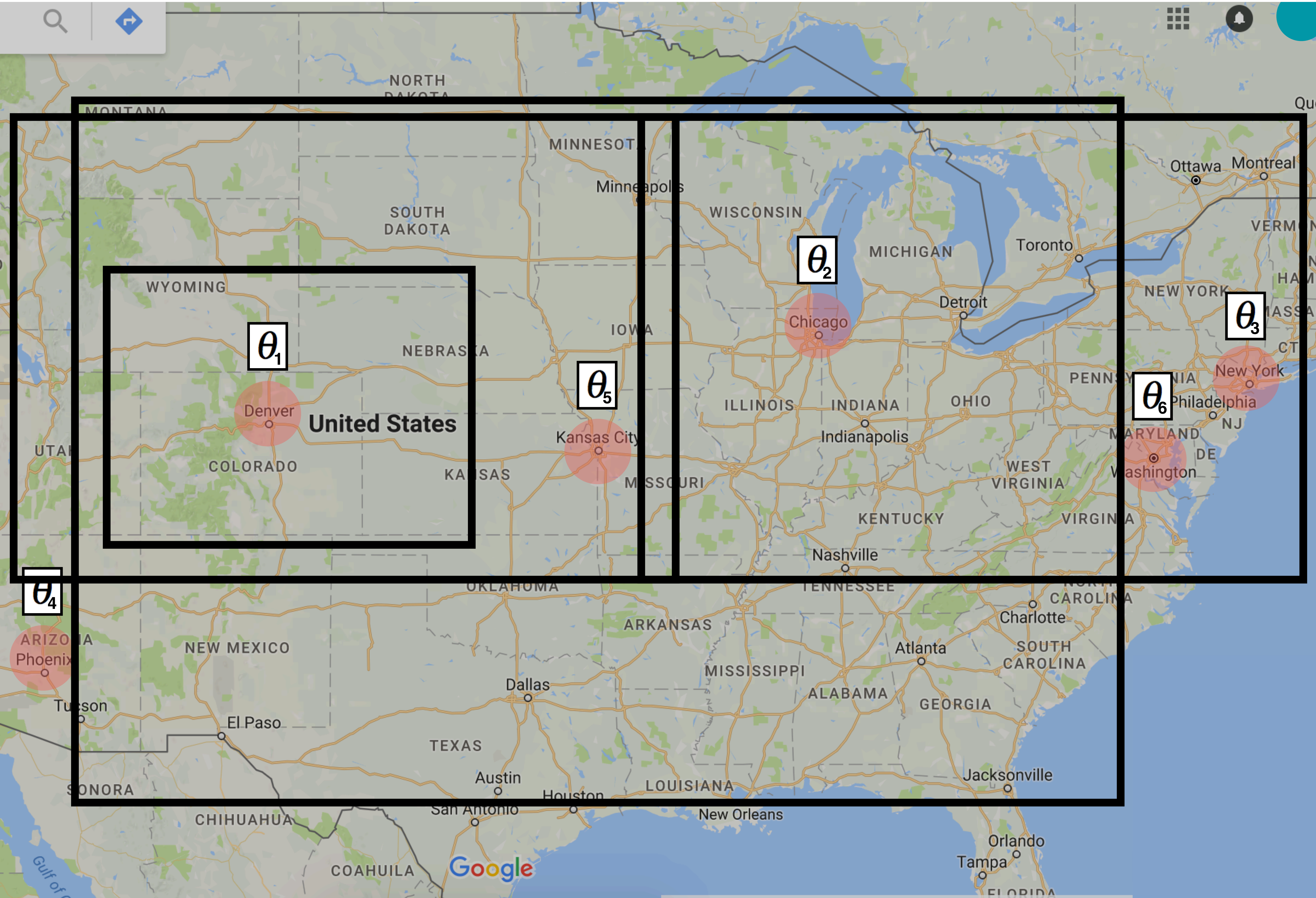
* The computer's Uncertainty about the user's goal

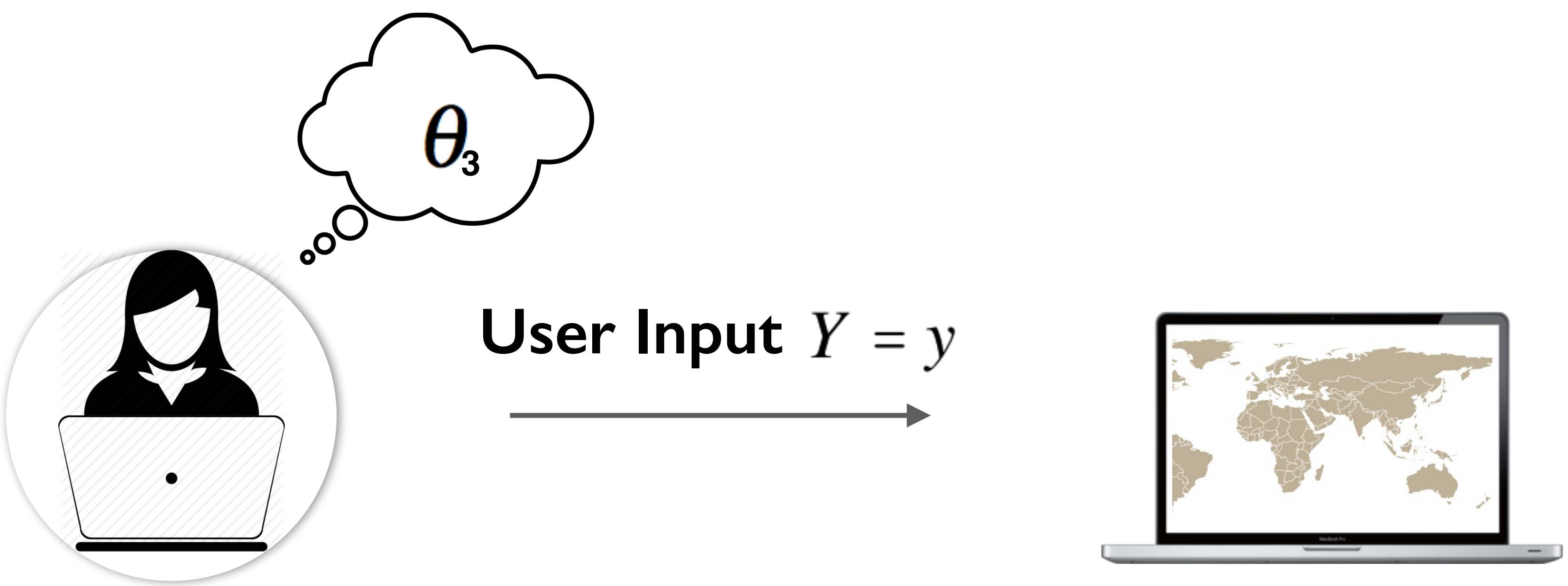
$$H(\Theta) = -\sum_{i=1}^{n} P(\Theta = \theta_i)\log_2 P(\Theta = \theta_i)$$

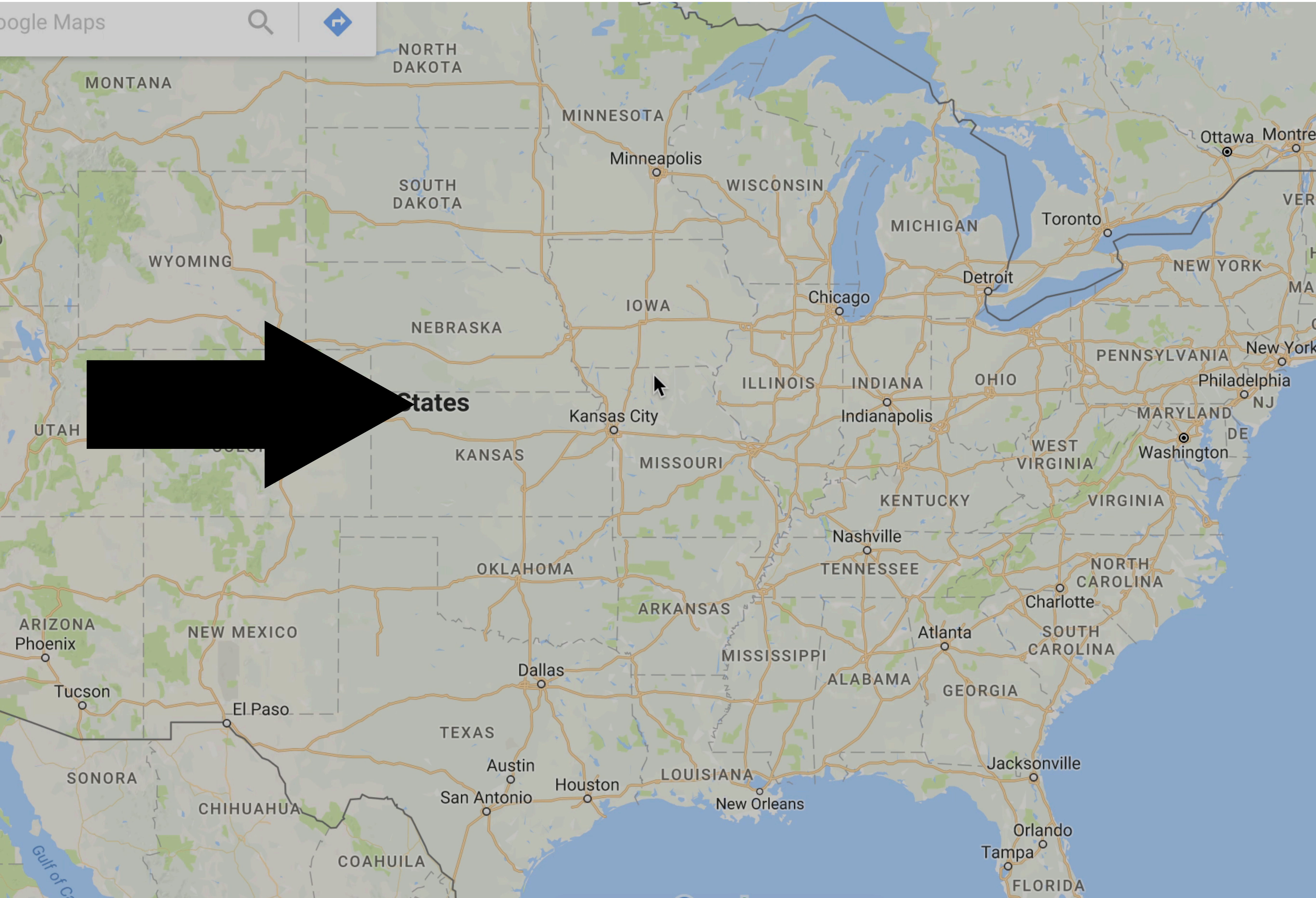

θ_3
View $X = x$

1948 1950 2017 **Bayesian Information Gain**

BIGnav




θ_3
User Input $Y = y$


oogle Maps
NORTH DAKOTA
MONTANA
MINNESOTA
Minneapolis
SOUTH DAKOTA
WISCONSIN
MICHIGAN
Toronto
Ottawa
Montre
WYOMING
NEW YORK
Detroit
Chicago
IOWA
NEBRASKA
PENNSYLVANIA
New York
States
ILLINOIS
INDIANA
OHIO
Philadelphia
NJ
UTAH
Kansas City
Indianapolis
MARYLAND
DE
KANSAS
MISSOURI
WEST VIRGINIA
Washington
KENTUCKY
VIRGINIA
Nashville
OKLAHOMA
TENNESSEE
NORTH CAROLINA
Charlotte
ARKANSAS
ARIZONA
Phoenix
NEW MEXICO
Atlanta
SOUTH CAROLINA
MISSISSIPPI
Dallas
ALABAMA
GEORGIA
Tucson
El Paso
TEXAS
SONORA
Austin
LOUISIANA
Jacksonville
Houston
San Antonio
New Orleans
CHIHUAHUA
Orlando
Tampa
COAHUILA
FLORIDA

View $X = x$

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

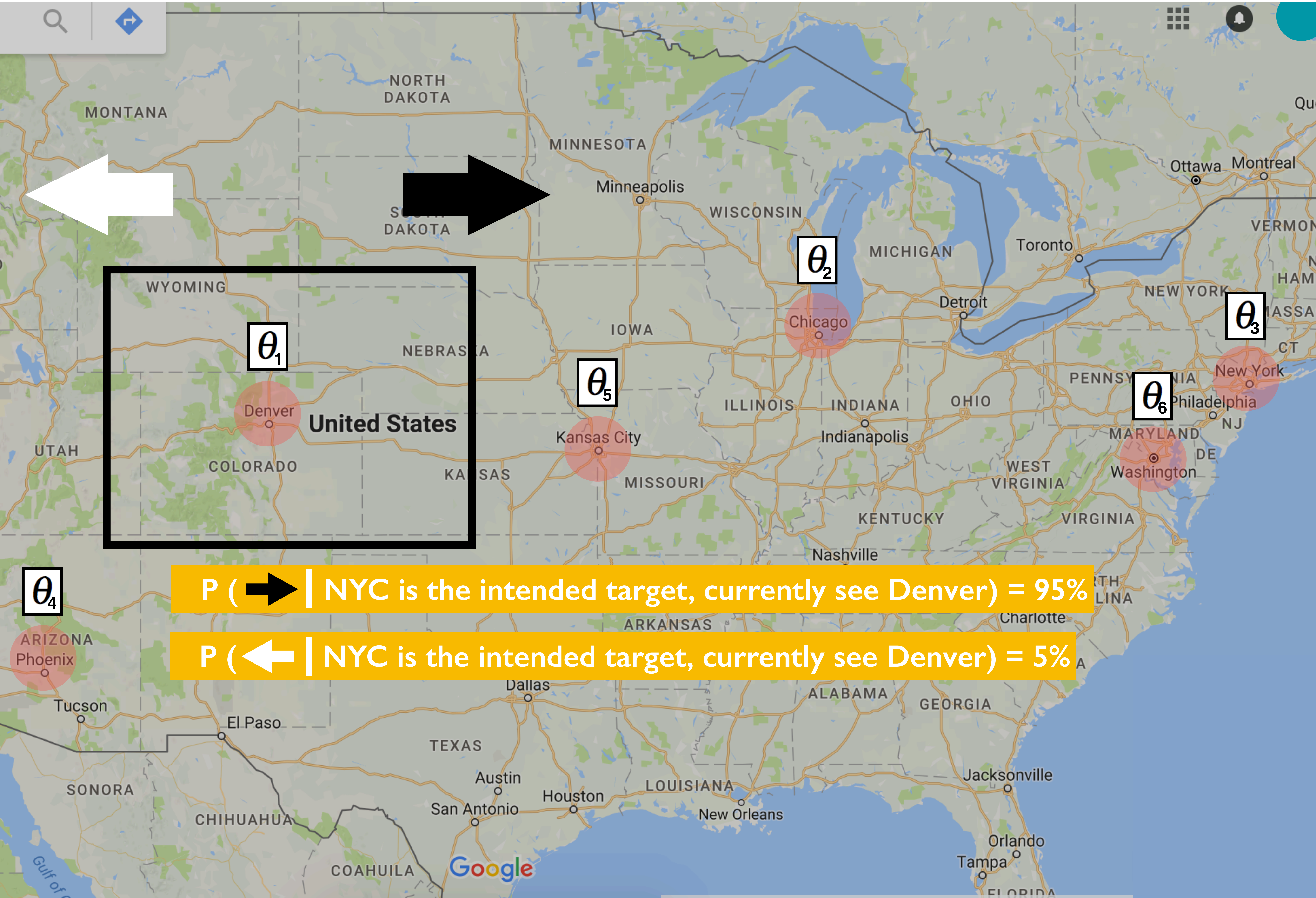
Interpret User Input

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

User Input $Y = y$

1948 — 1950 — 2017 **Bayesian Information Gain**

BIGnav



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

Update its knowledge

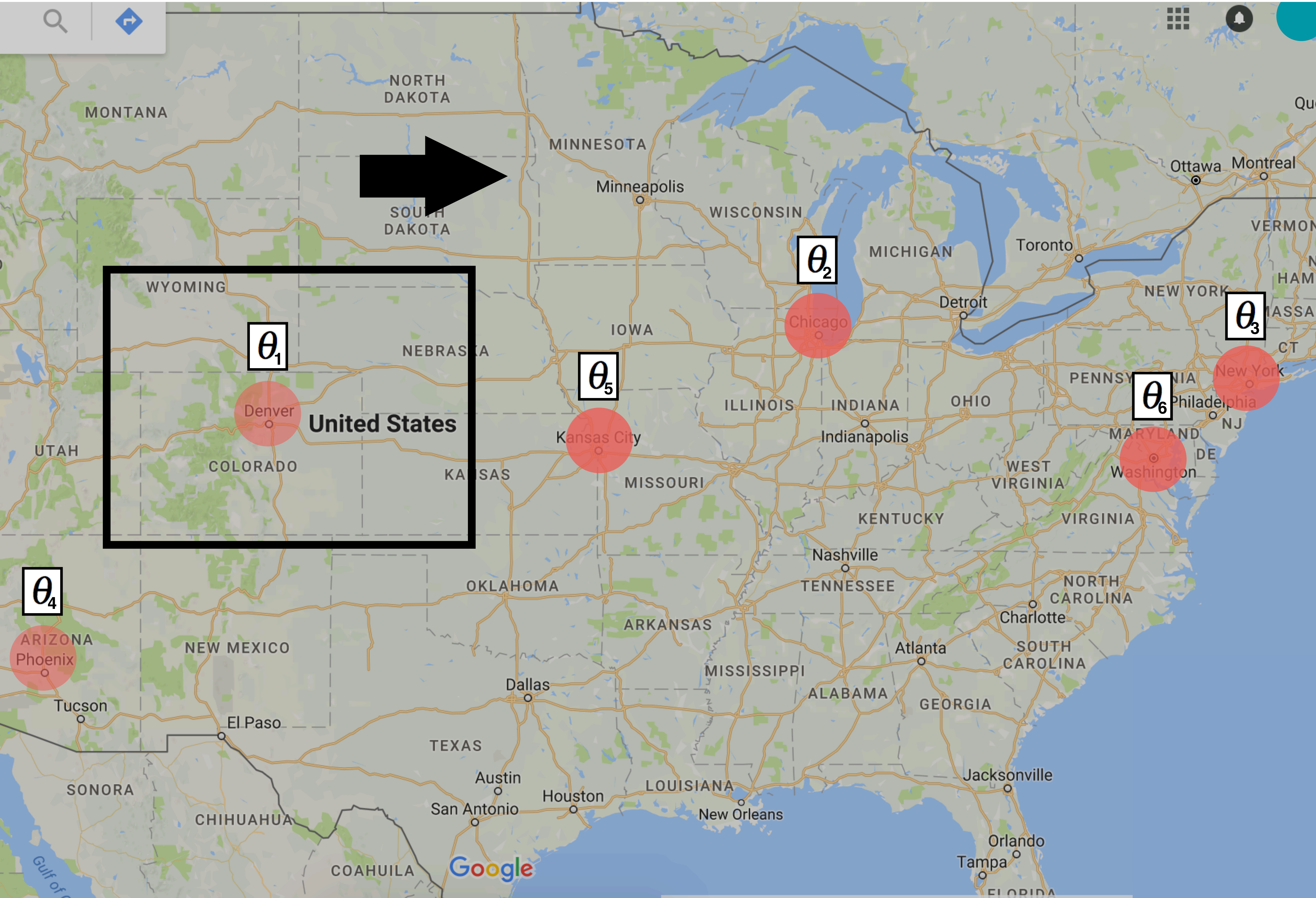
View $X = x$

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

User Input $Y = y$

1948 1950 2017 **Bayesian Information Gain**

BIGnav



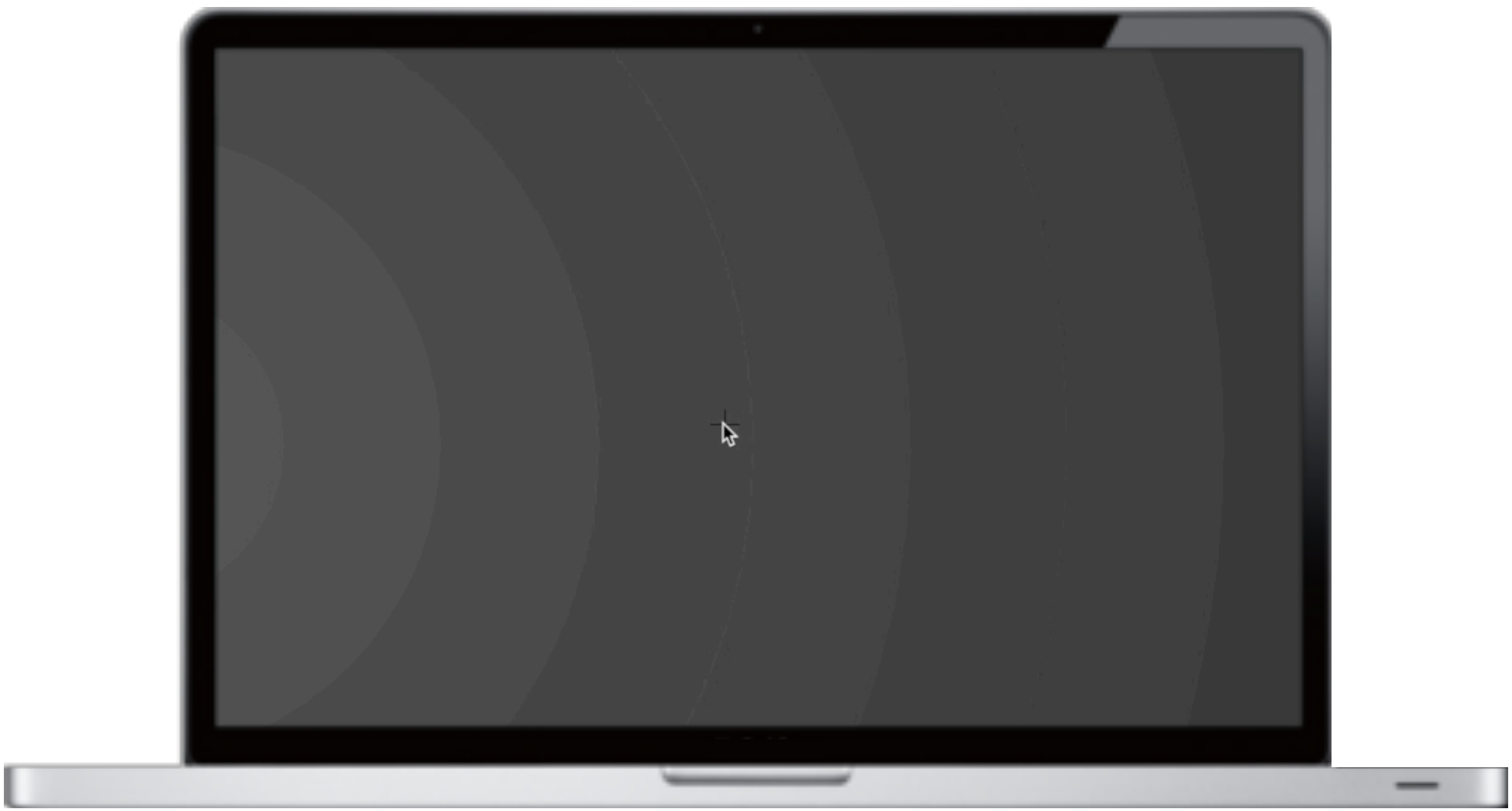
- The computer's Uncertainty about the user's goal

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

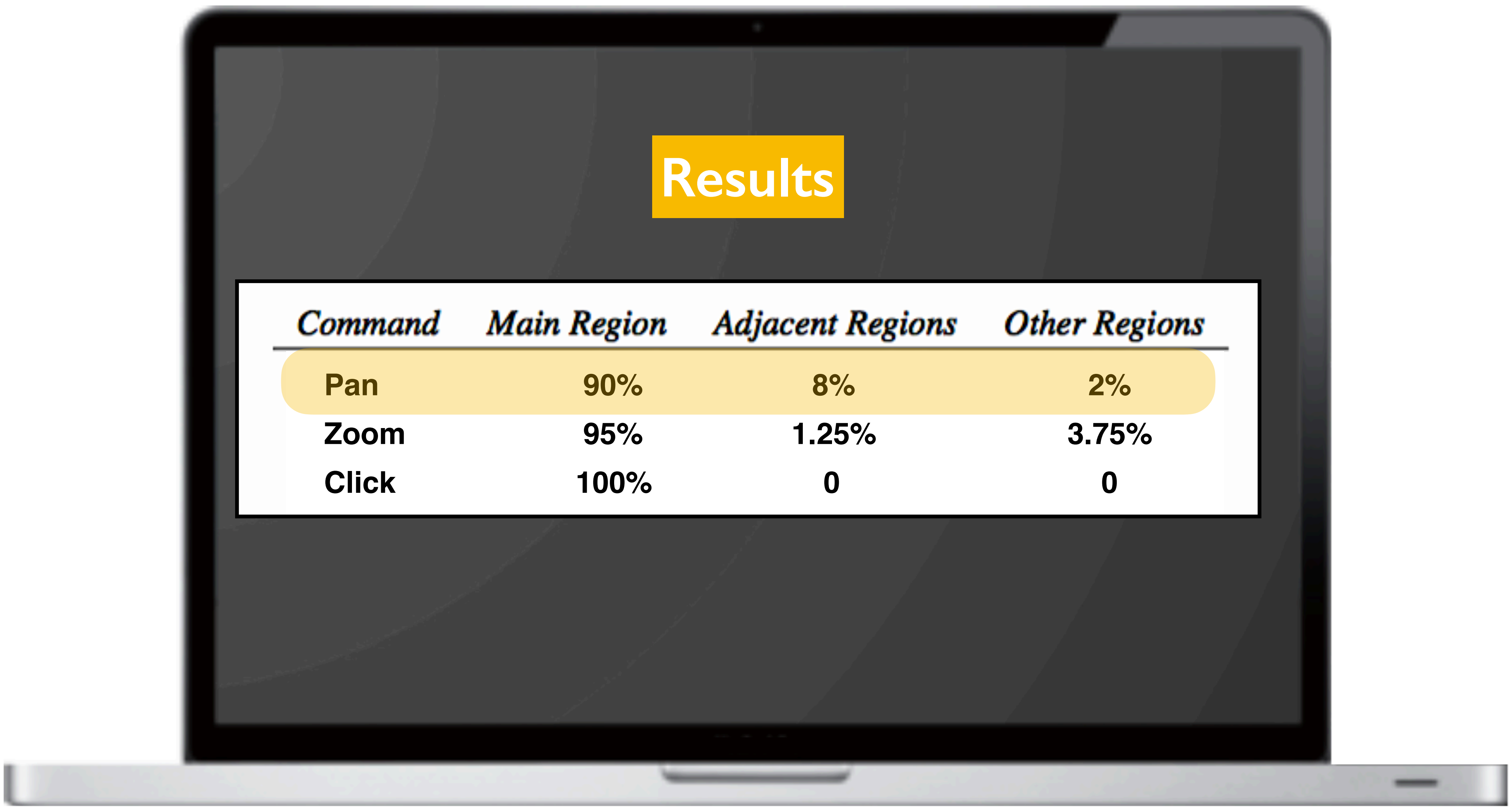
- The computer's updated knowledge about the user's goal

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

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



BIGnav



Results

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

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

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

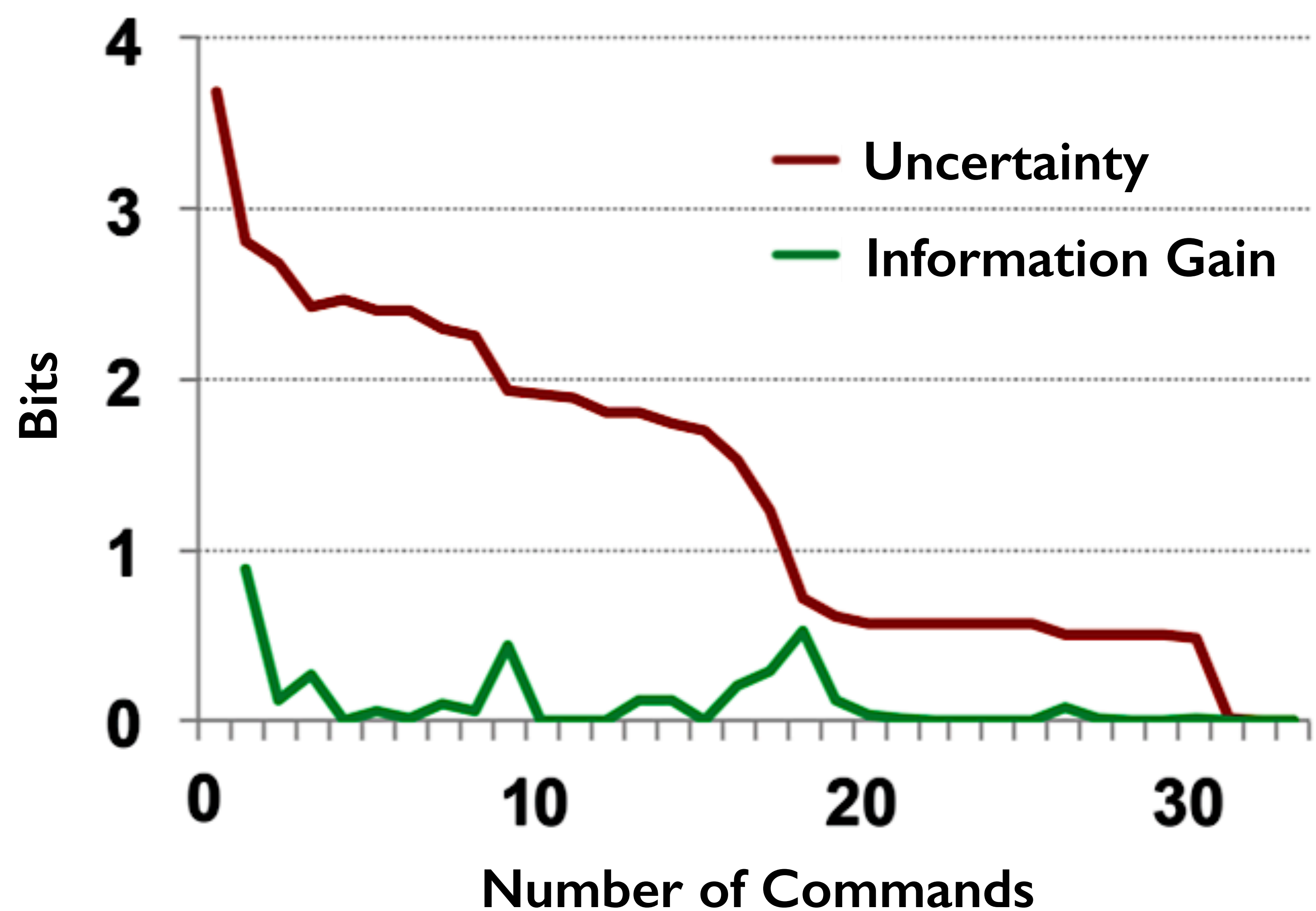
- The computer's updated knowledge about the user's goal

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

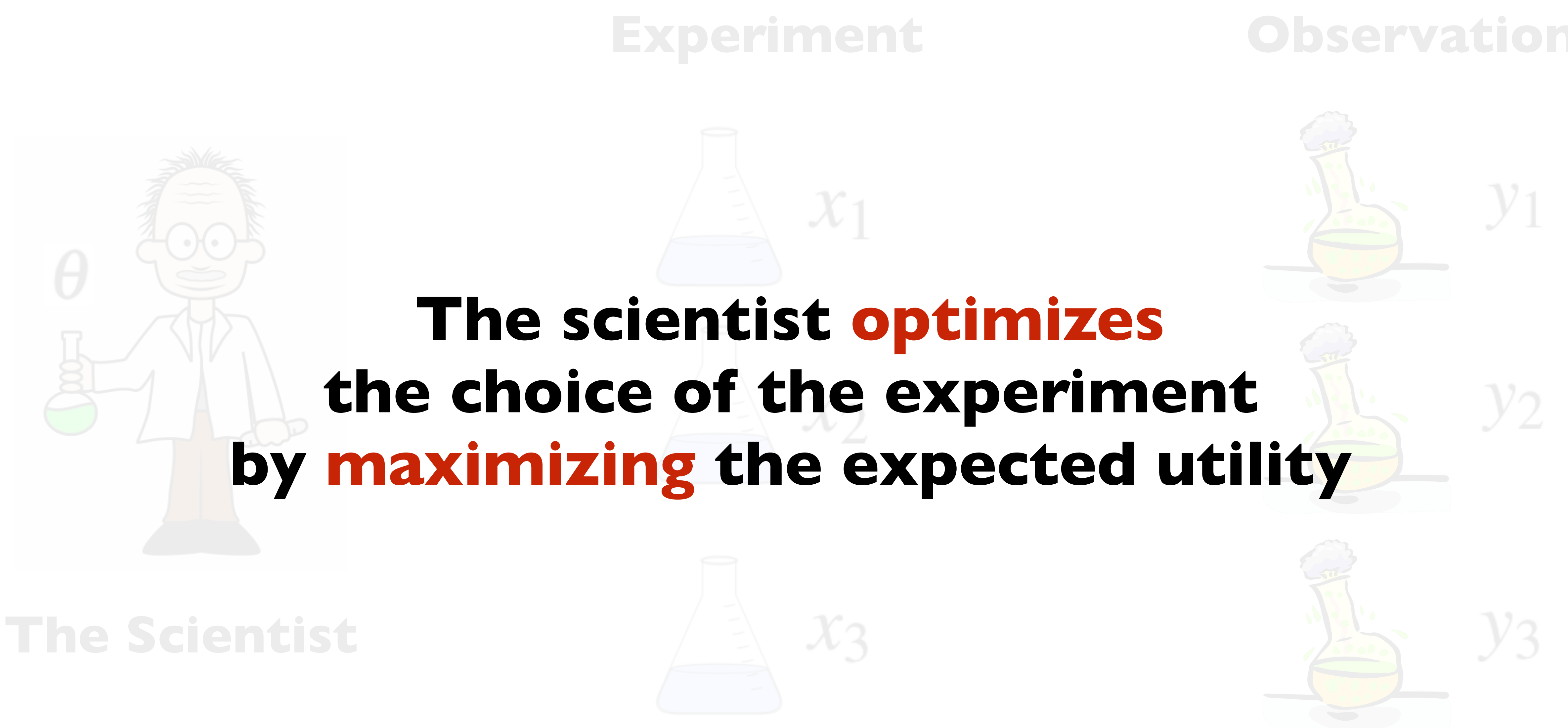
- The information in the user's input for reducing the computer's uncertainty

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

- Executes the user input only Multiscale navigation
- Each user input does not provide much information for the computer to know her goal

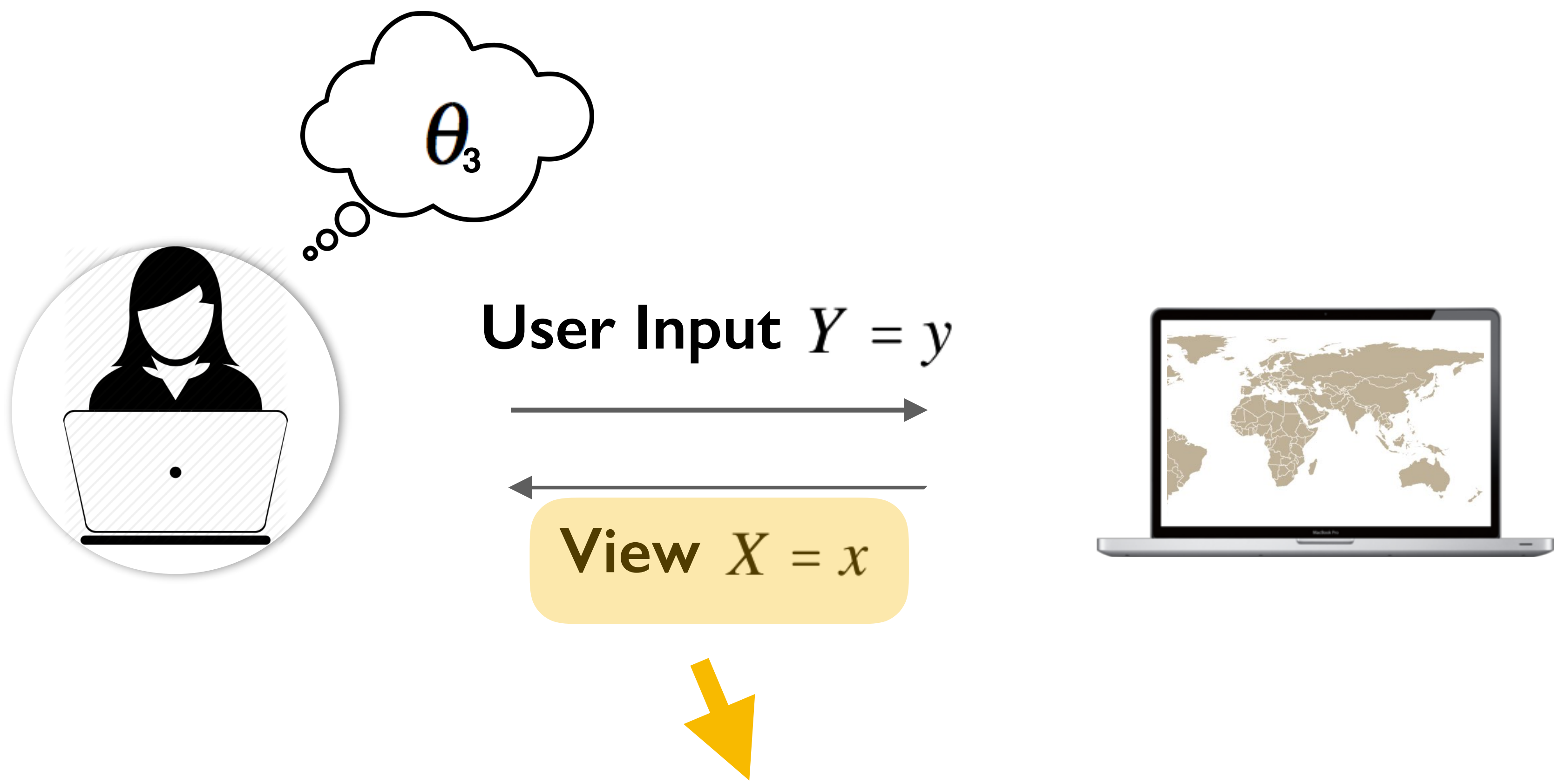


Can we challenge users to give more information?



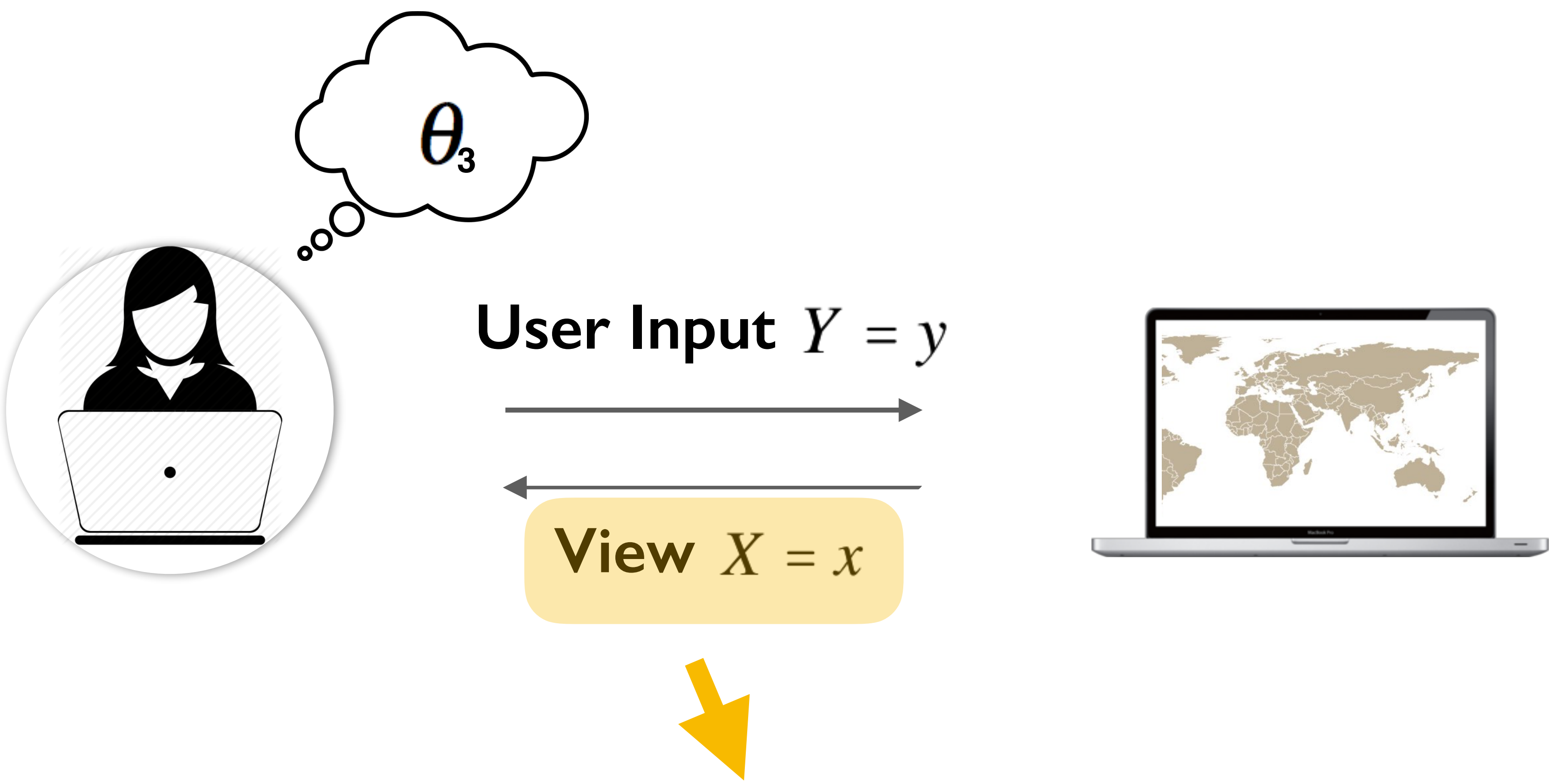
**The scientist optimizes
the choice of the experiment
by maximizing the expected utility**

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



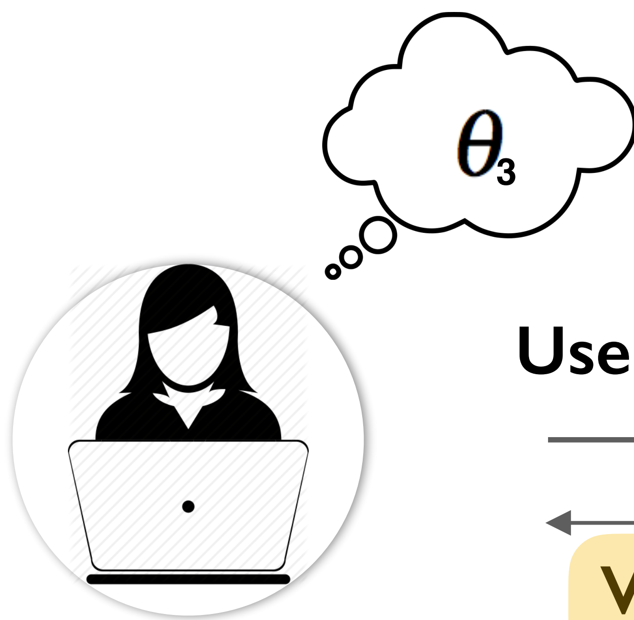
The expected information gain

Wanyu Liu, Rafael Lucas D'Oliveira, Michel Beaudouin-Lafon, and Olivier Rioul. BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. (CHI '17).



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

Wanyu Liu, Rafael Lucas D'Oliveira, Michel Beaudouin-Lafon, and Olivier Rioul. BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. (CHI '17).



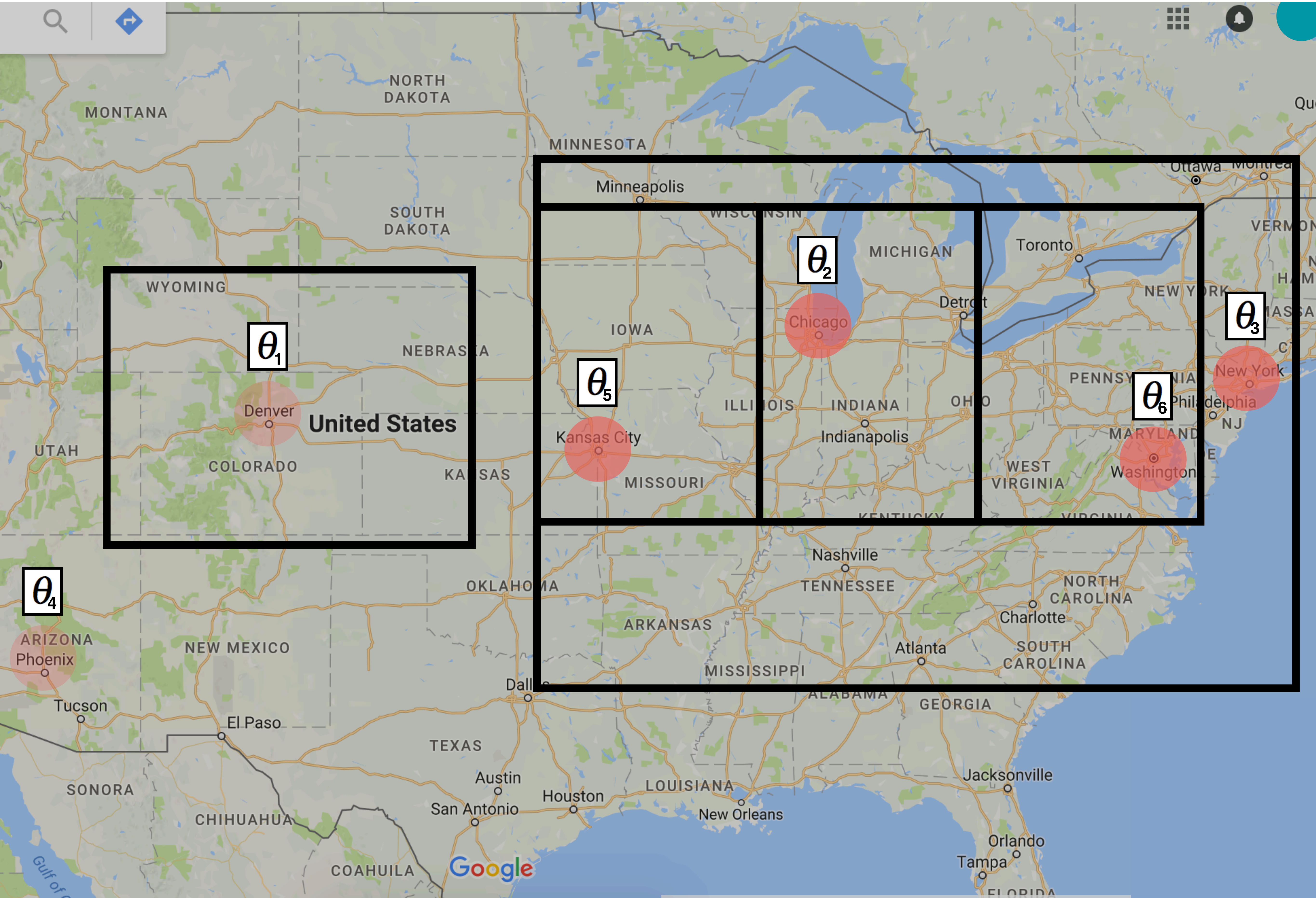
User Input $Y = y$

View $X = x$

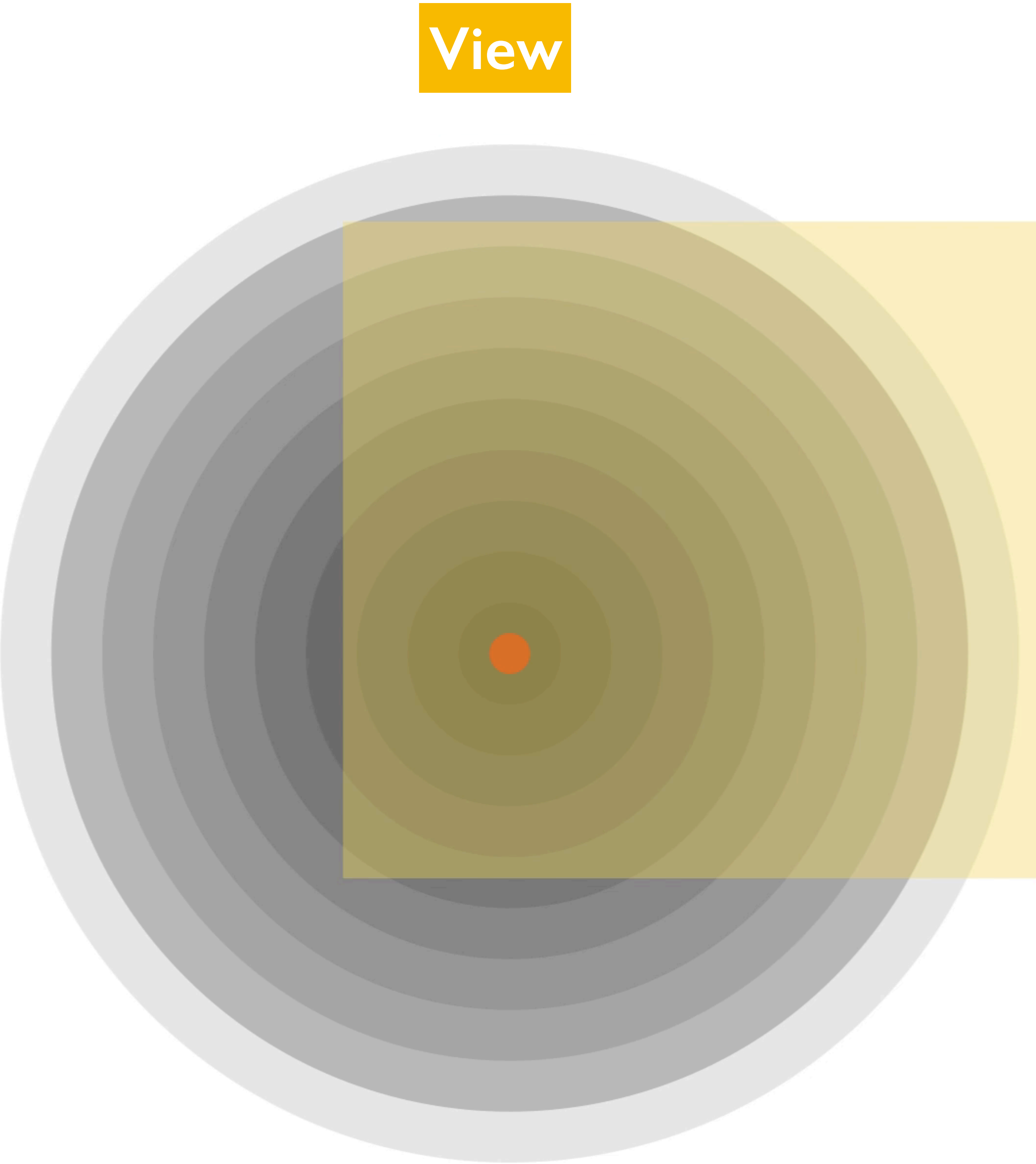
Go over all possible feedback, and find the one that maximizes the expected information gain

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

Wanyu Liu, Rafael Lucas D'Oliveira, Michel Beaudouin-Lafon, and Olivier Rioul. BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. (CHI '17).

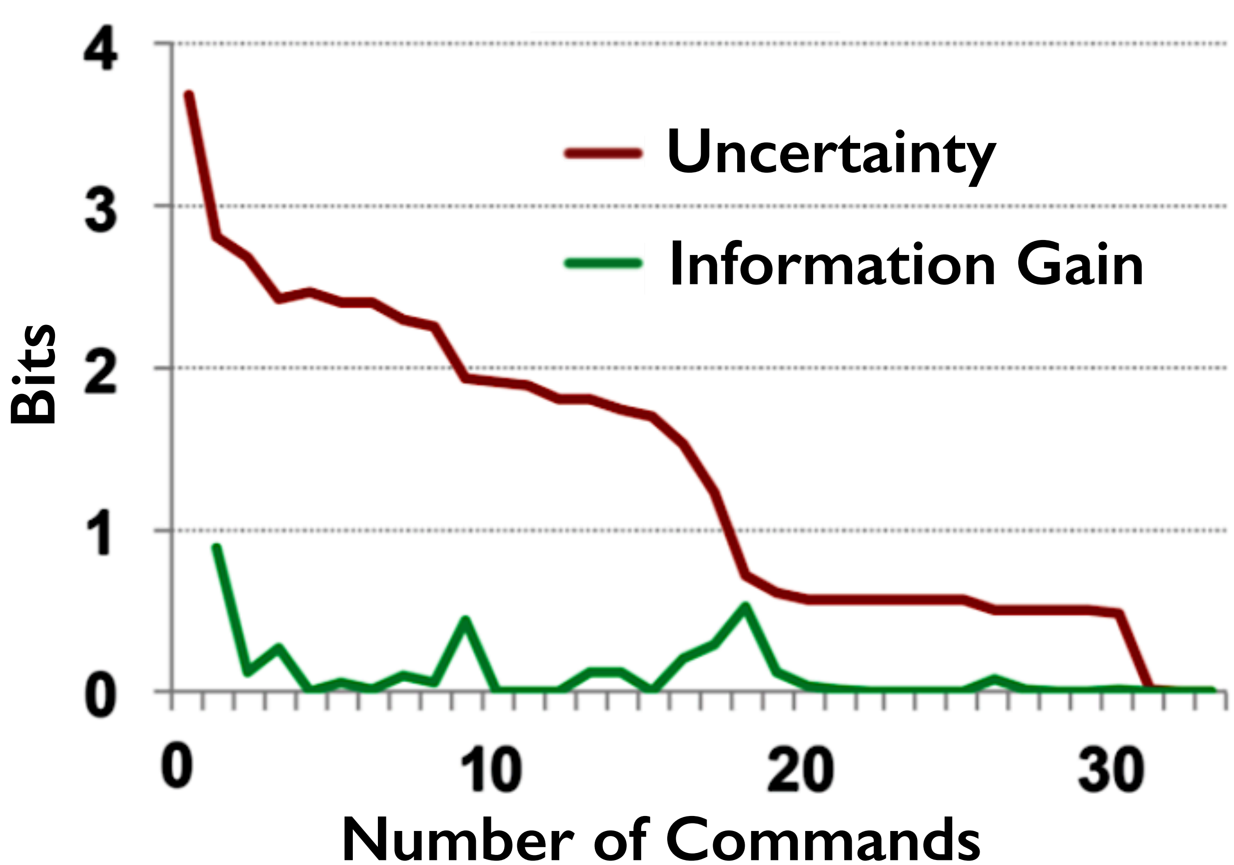

θ_1
θ_2
θ_3
θ_4
θ_5
θ_6
Denver
United States
Kansas City
Chicago
New York
Washington
Phoenix
Minneapolis
Toronto
Detroit
Indianapolis
Nashville
Atlanta
Charlotte
Tucson
El Paso
Austin
San Antonio
Houston
New Orleans
Jacksonville
Orlando
Tampa
Ottawa
MONTANA
NORTH DAKOTA
SOUTH DAKOTA
MINNESOTA
WYOMING
NEBRASKA
IOWA
MICHIGAN
INDIANA
MISSOURI
COLORADO
UTAH
ARIZONA
NEW MEXICO
OKLAHOMA
TENNESSEE
ARKANSAS
MISSISSIPPI
GEORGIA
NORTH CAROLINA
SOUTH CAROLINA
WEST VIRGINIA
TEXAS
LOUISIANA
SONORA
CHIHUAHUA
COAHUILA
NJ
Google

BIGnav



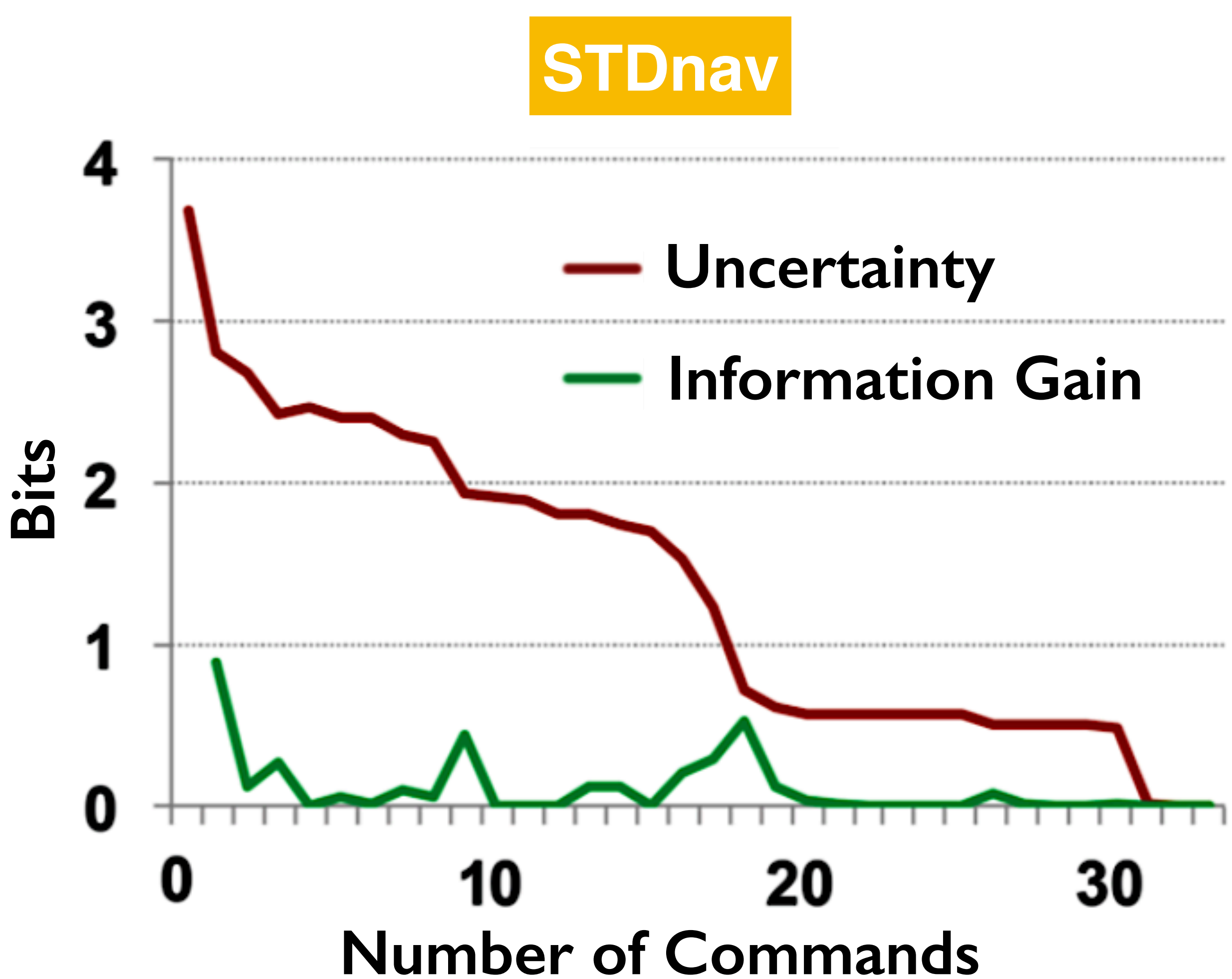
* BIGnav gains maximum information from each user input

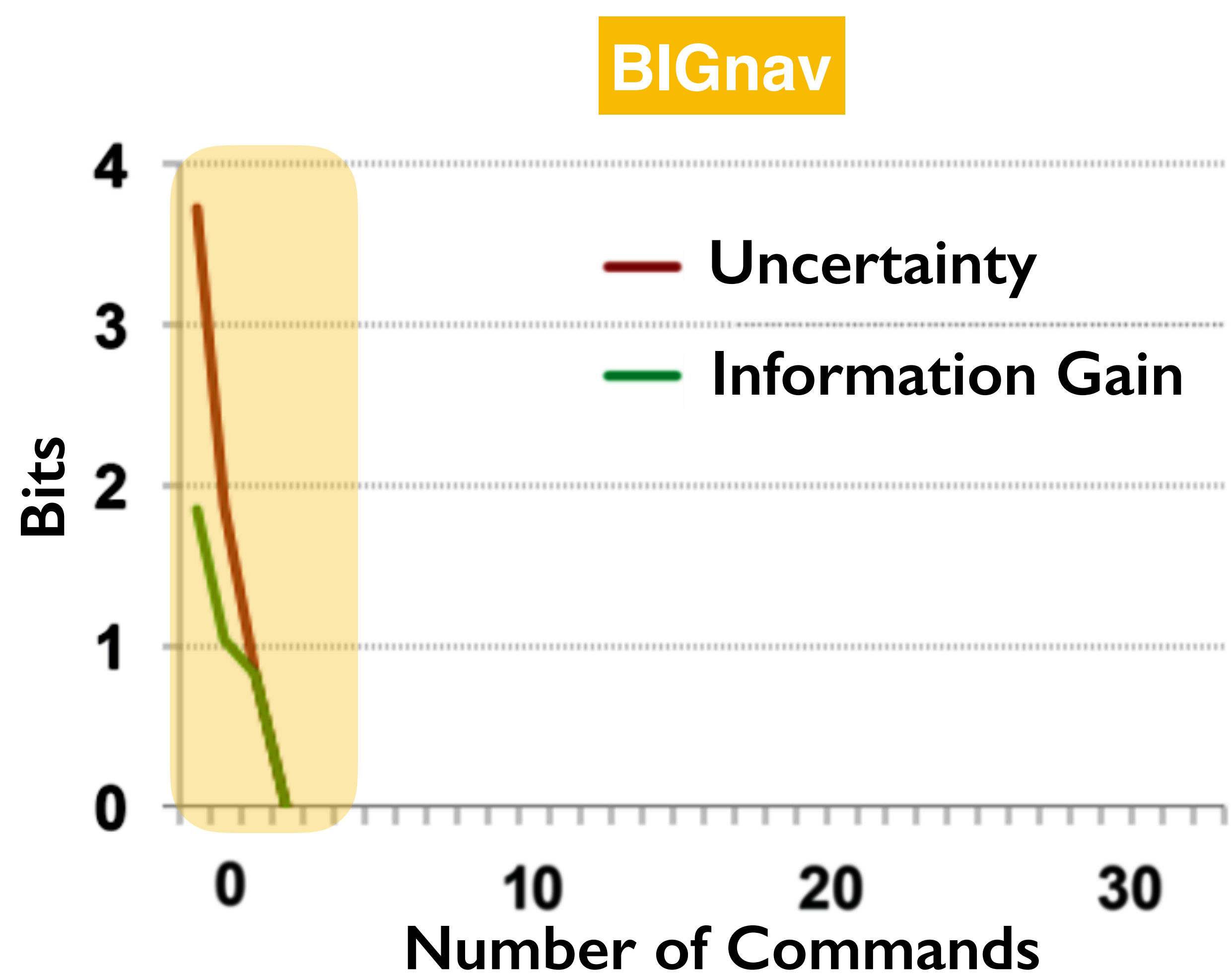
STDnav



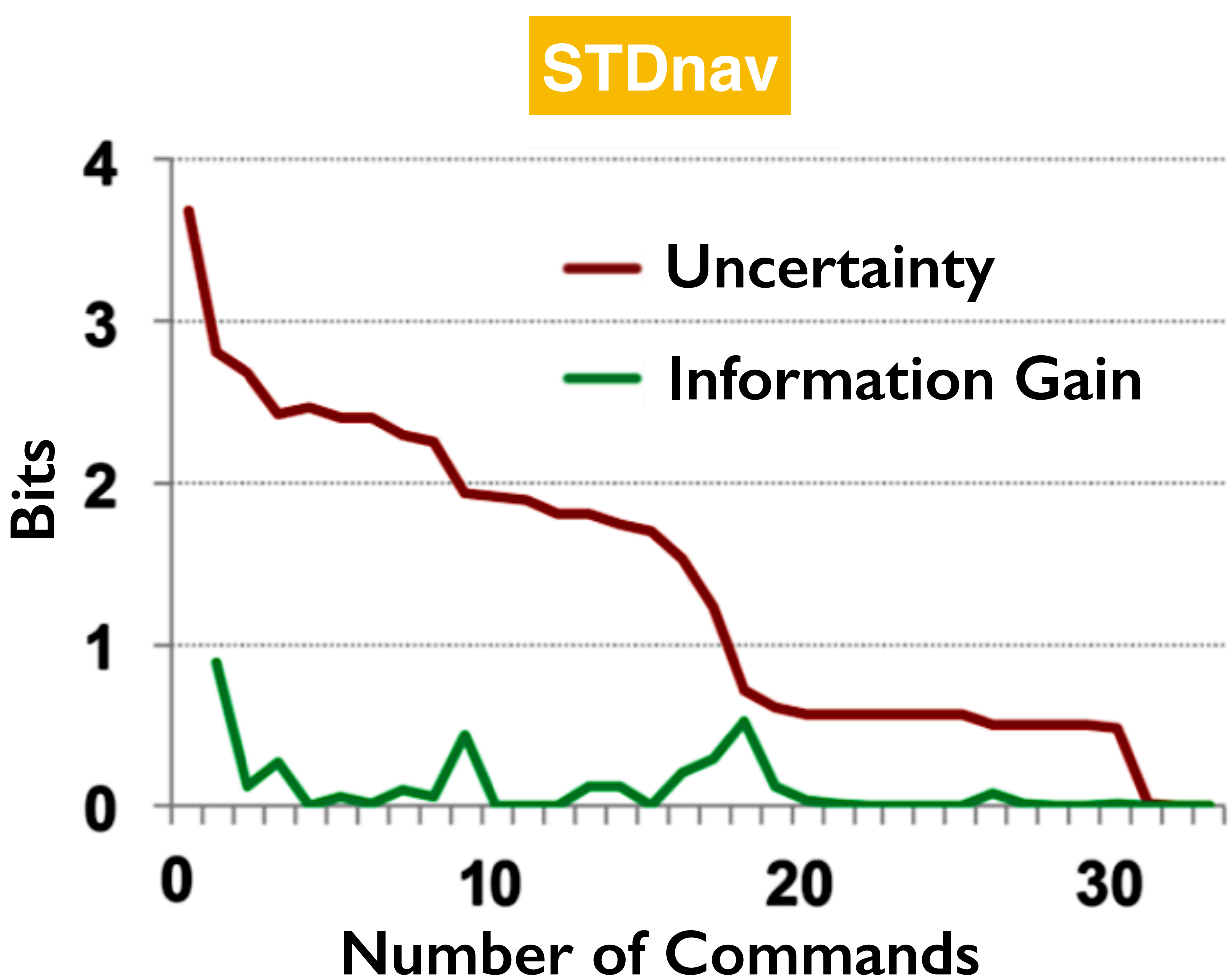
BIGnav

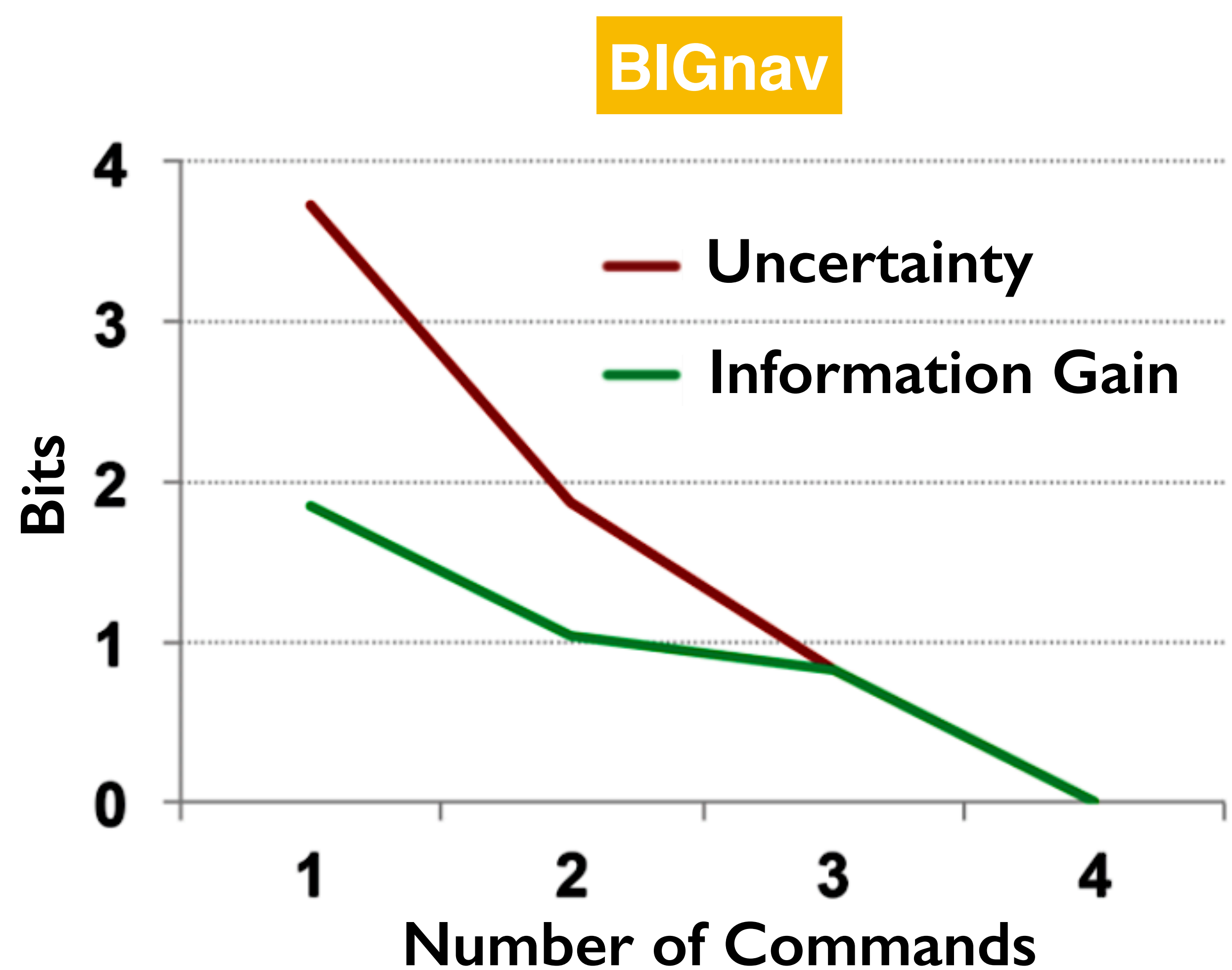
- BIGnav gains maximum information from each user input





* BIGnav gains maximum information from each user input





BIGmap

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.

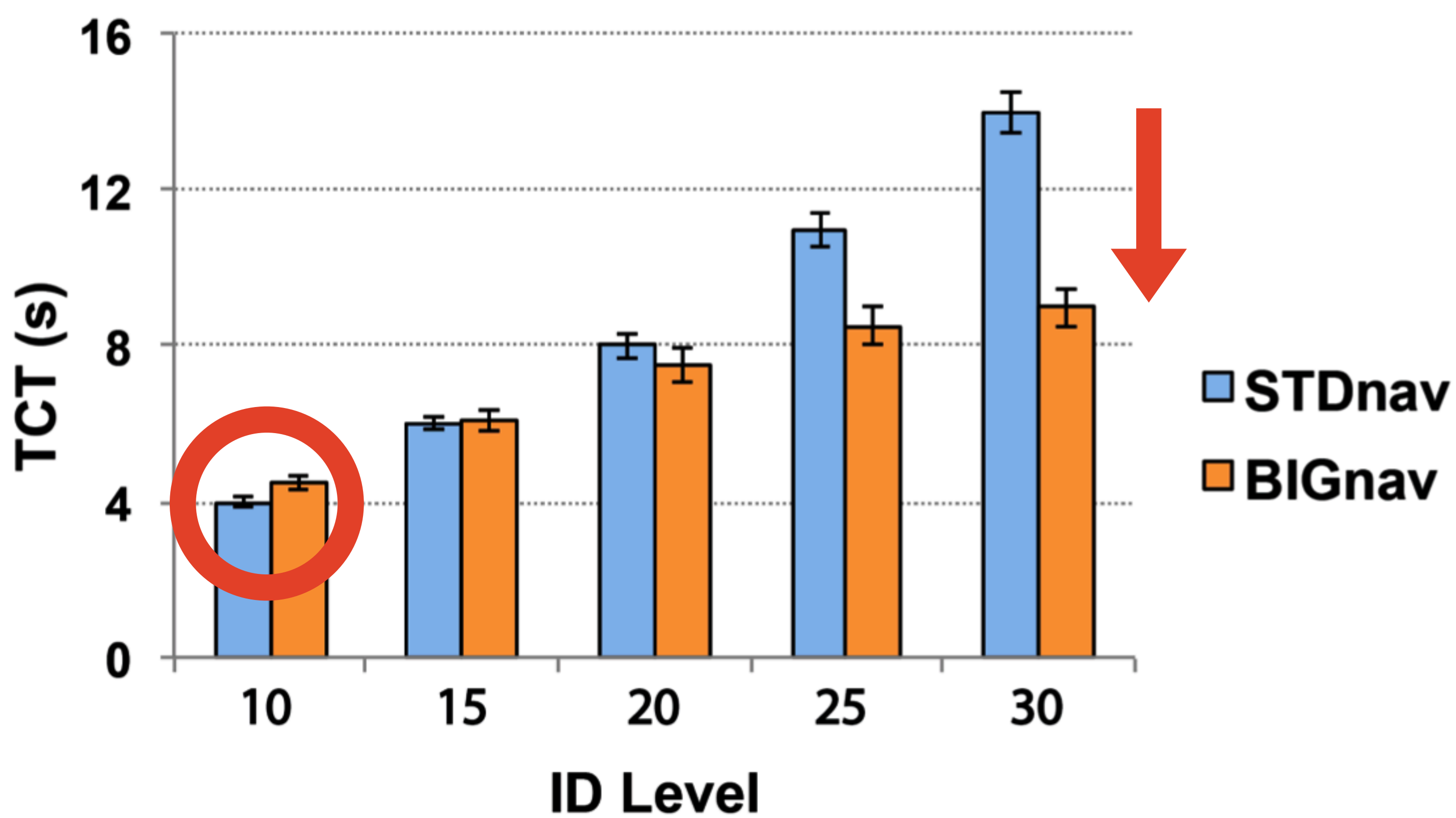
- Full factorial within-participant design:

16 Participant
x 2 Navigation Technique
x 5 Index of Difficulty x 6 Distribution
x 5 Replication

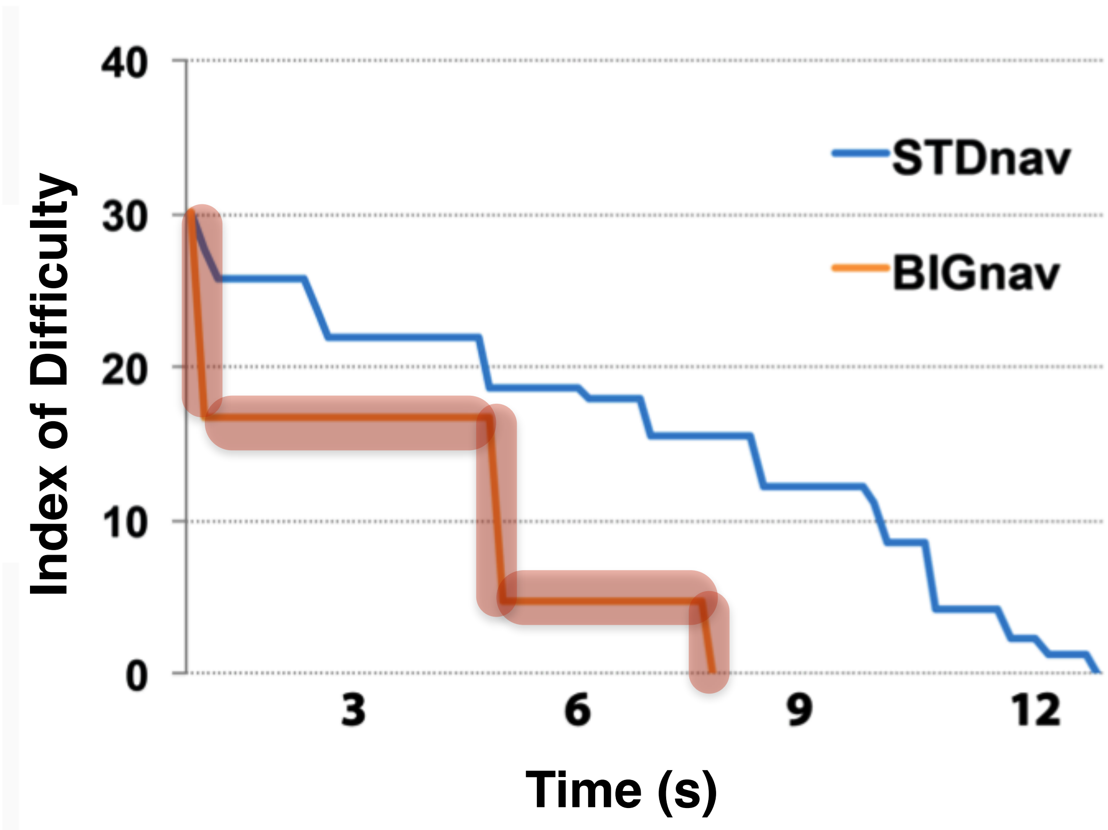
- Technique: BIGnav, STDnav

- Index of Difficulty: 10, 15, 20, 25, 30

* The further the target is located, the better BIGnav performs



* Trajectory in multiscale worlds.
Though being more efficient, BIGnav incurs a higher cognitive load



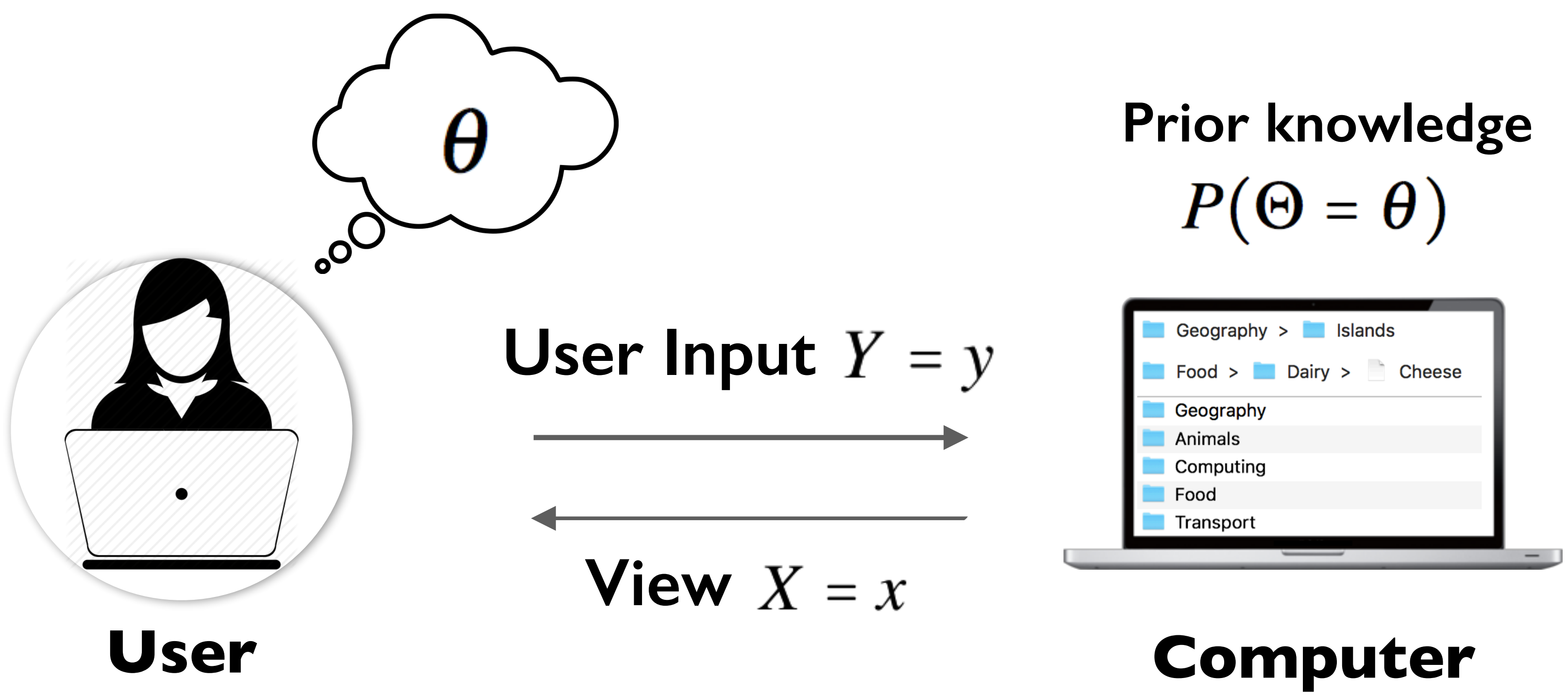
Experiment Summary

BIGnav is up to 40% faster than STDnav

Half of the participants preferred BIGnav for being efficient and interactive

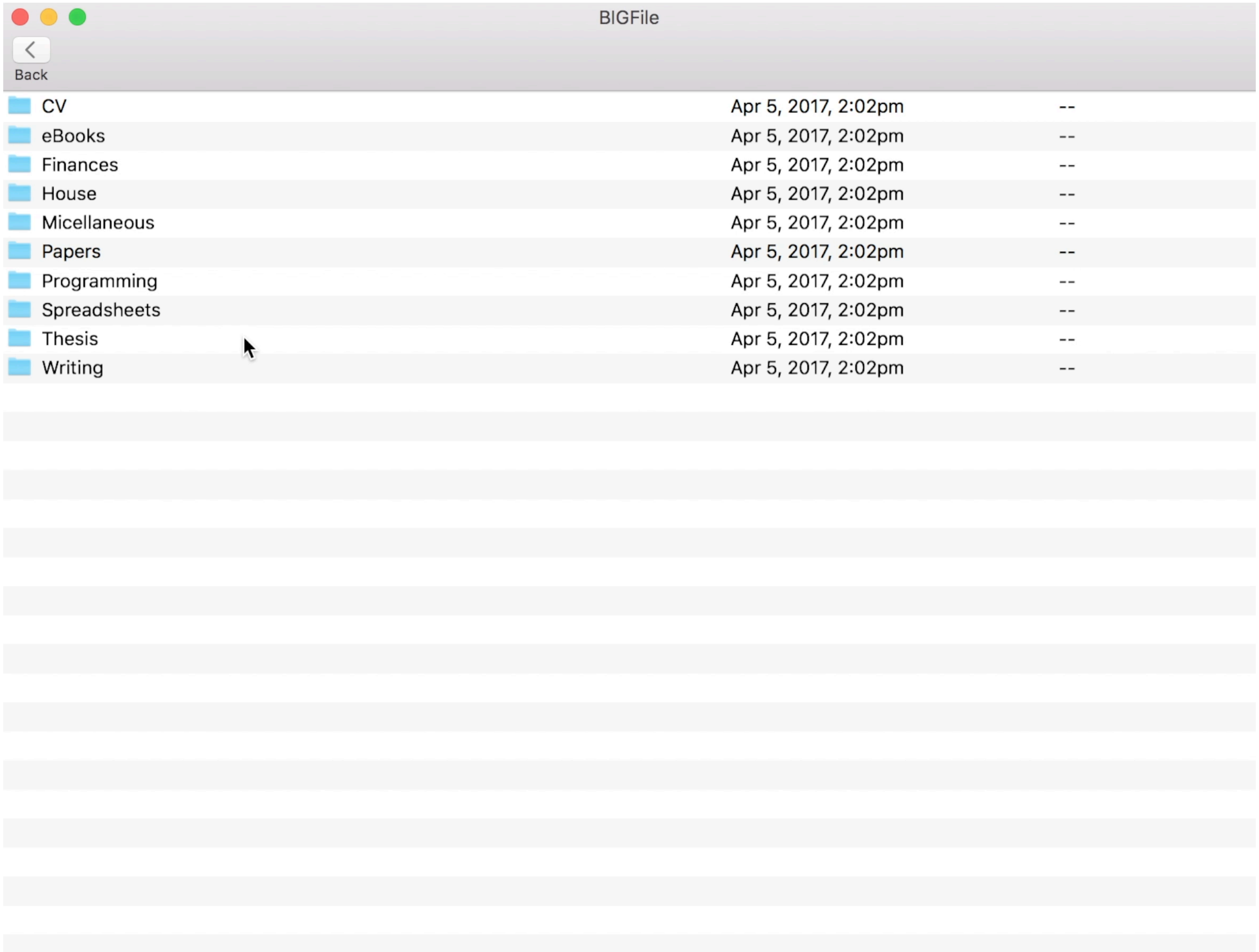
The other half favored STDnav for being comfortable and intuitive

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



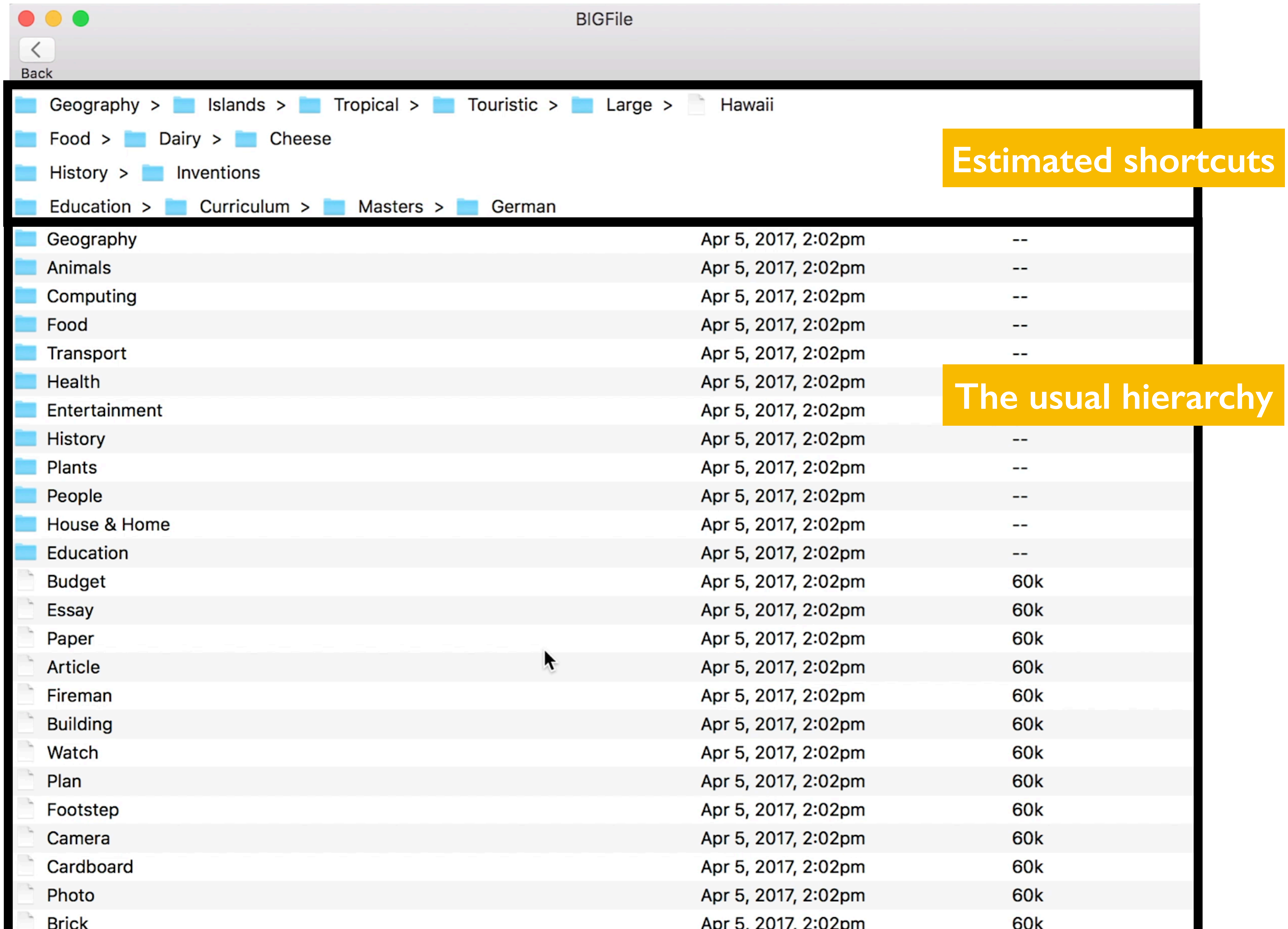
Wanyu Liu, Olivier Rioul, Joanna McGrenere, Wendy Mackay, and Michel Beaudouin-Lafon. BIGFile: Bayesian Information Gain for Fast File Retrieval. (CHI '18).

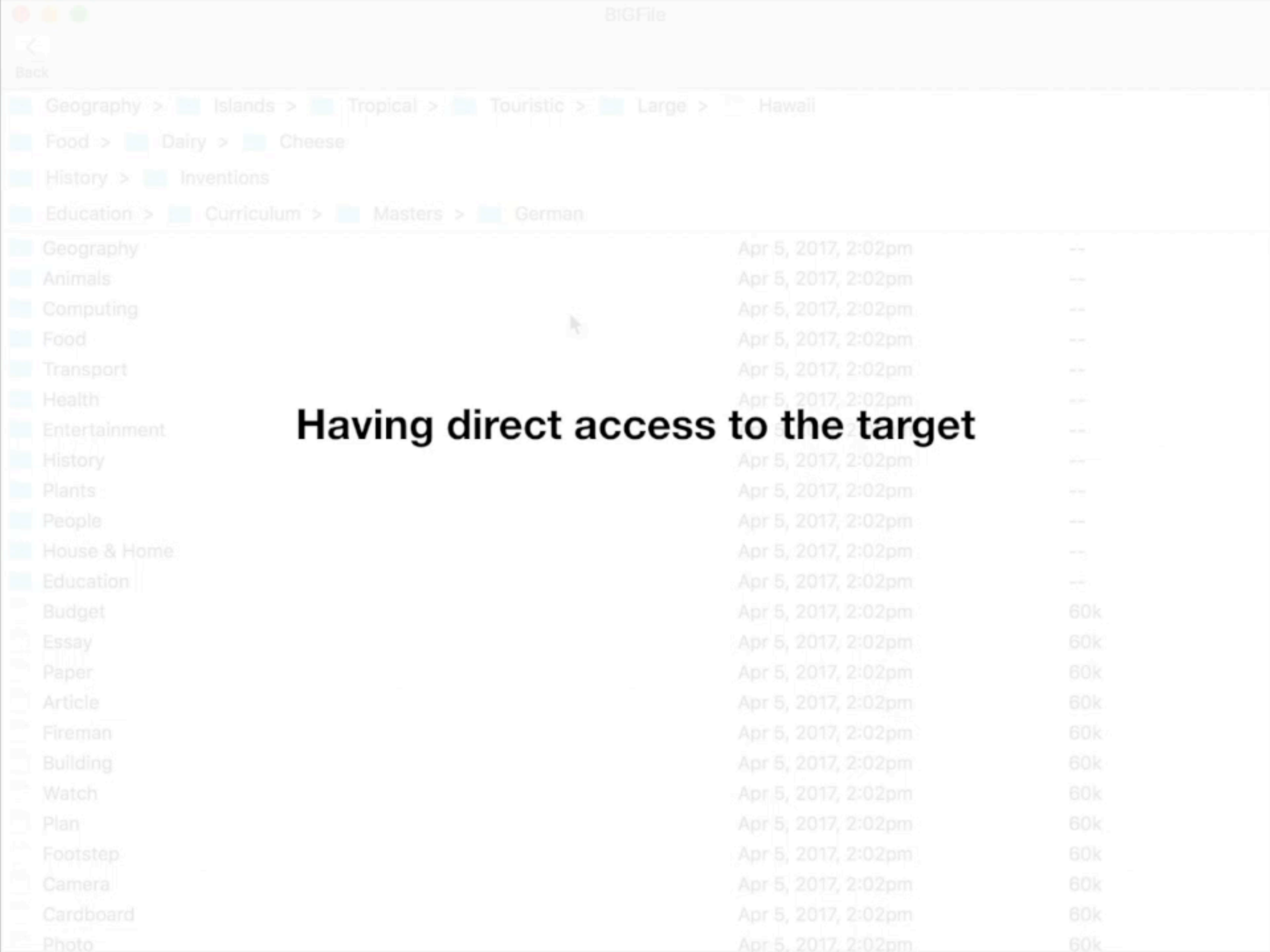
BIGFile


BIGFile
Back
CV Apr 5, 2017, 2:02pm --
eBooks Apr 5, 2017, 2:02pm --
Finances Apr 5, 2017, 2:02pm --
House Apr 5, 2017, 2:02pm --
Micellaneous Apr 5, 2017, 2:02pm --
Papers Apr 5, 2017, 2:02pm --
Programming Apr 5, 2017, 2:02pm --
Spreadsheets Apr 5, 2017, 2:02pm --
Thesis Apr 5, 2017, 2:02pm --
Writing Apr 5, 2017, 2:02pm --

1948 1950 2017 **Bayesian Information Gain**

BIGFile





Having direct access to the target

BIGFile

* [3 x 2] within-participant design:

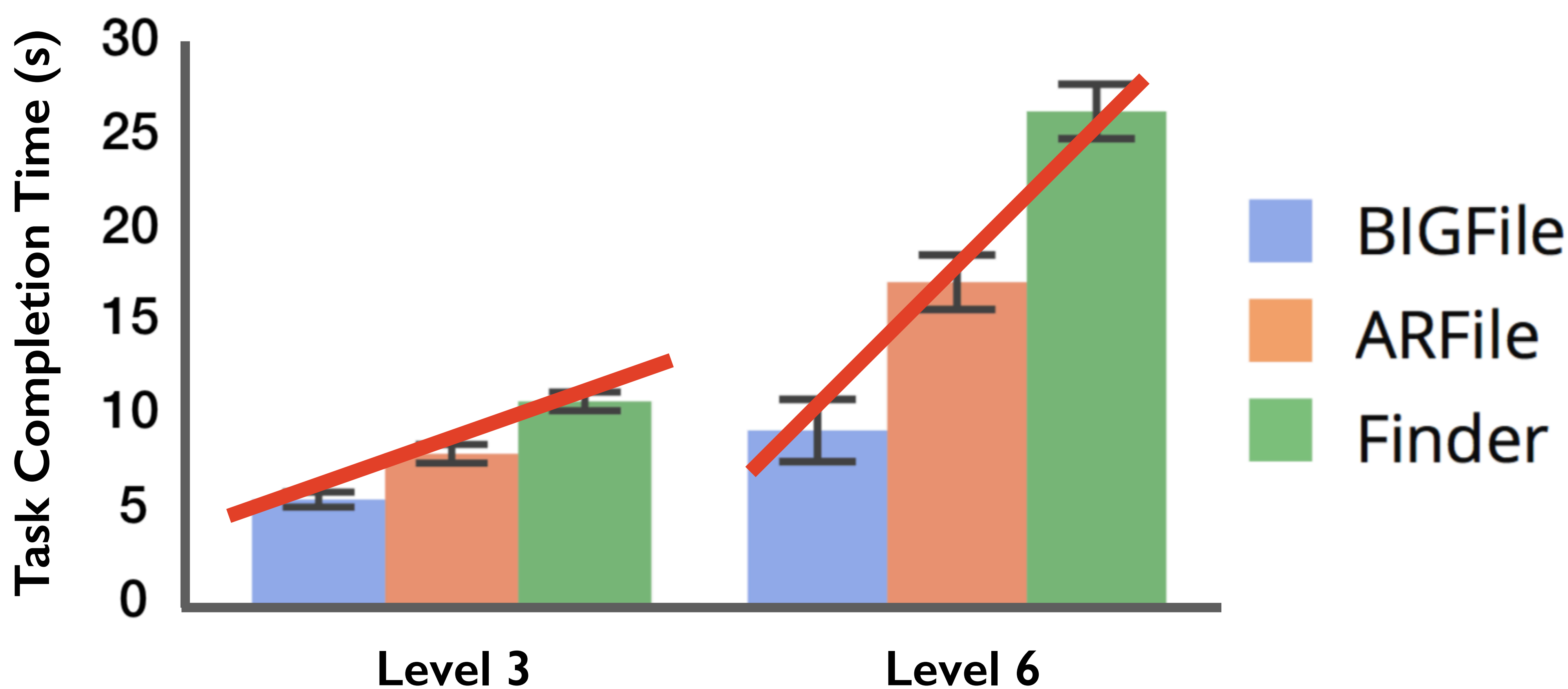
18 Participant
x 3 Interface
x 2 Target Level

* Interface: BIGFile ARFile Finder

* Target Level: 3, 6

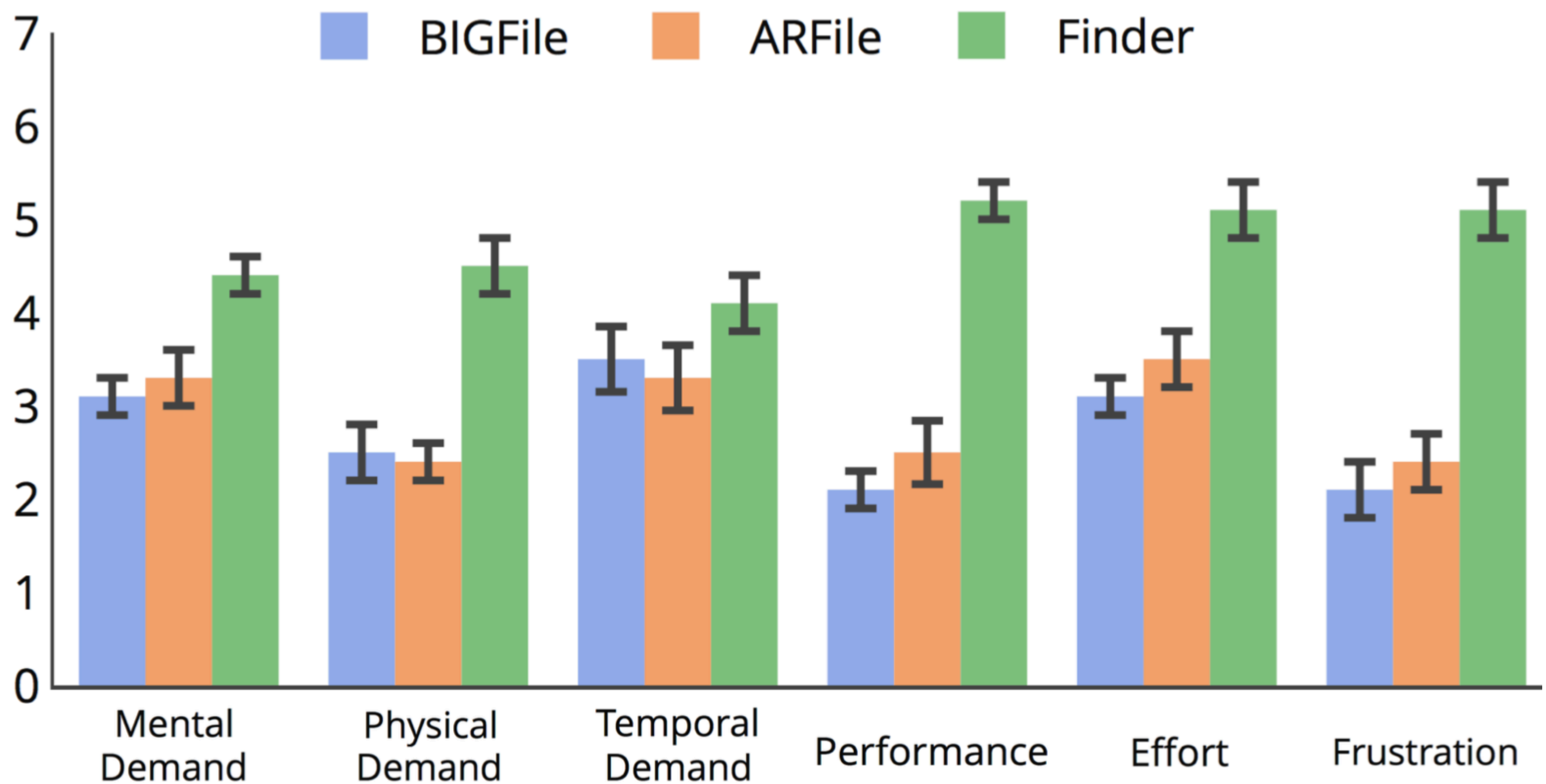
AccessRank: predicting what users will do next (Fitchett & Cockburn 2012)

* BIGFile saves time to retrieve a file



- Only data from the second session is shown here.

Both BIGFile and ARFile are preferred by participants



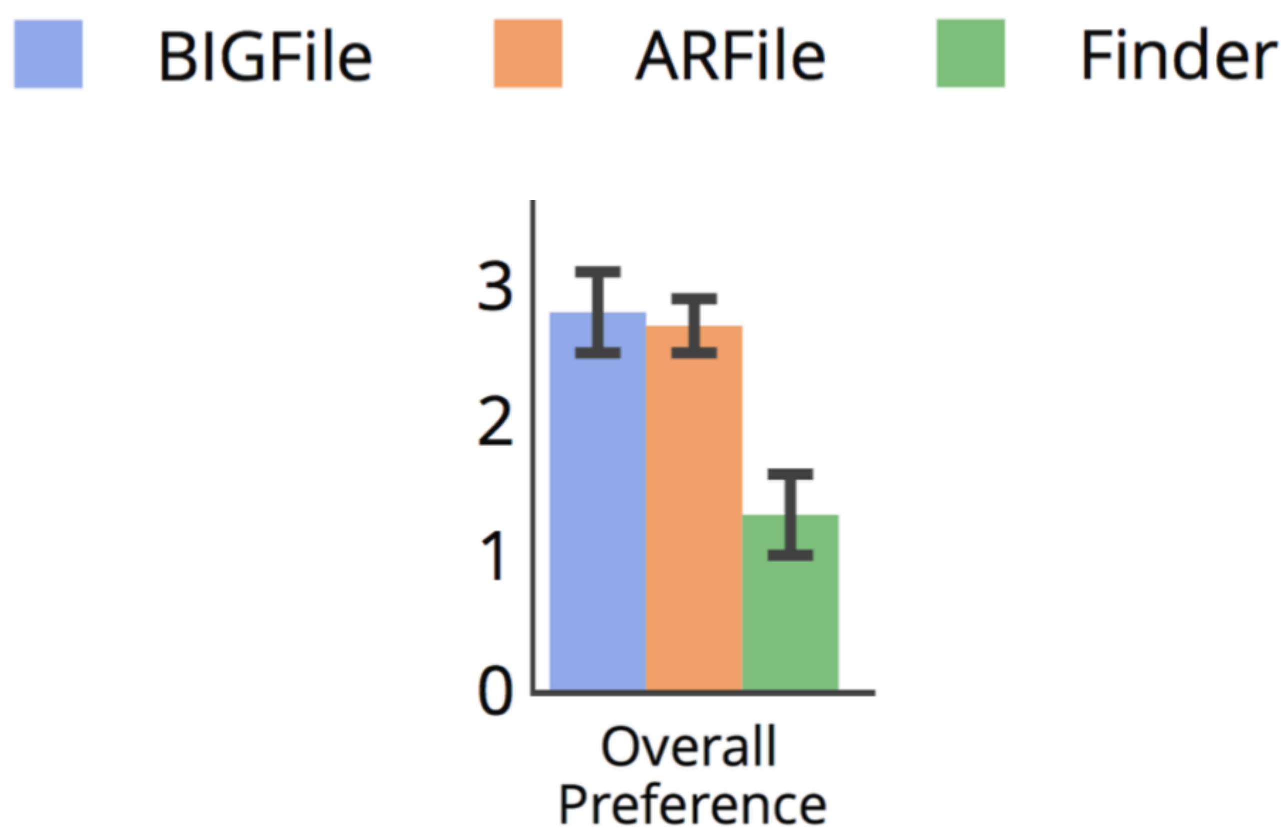
- NASA TLX scores: lower is better.

Experiment Summary

BIGFile is up to 44% faster than ARFile and 64% faster than Finder

Experiment Summary

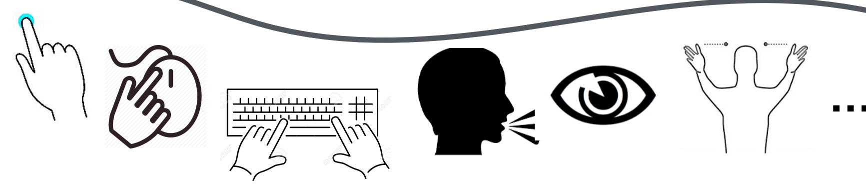
Both BIGFile and ARFile are preferred by participants



- Overall preference: higher is better.

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

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 $X = x$

Input

Interface

Time, errors

User

Computer

Drawback 1: Speed-accuracy tradeoff

Solution: Control errors

Solution: Control errors

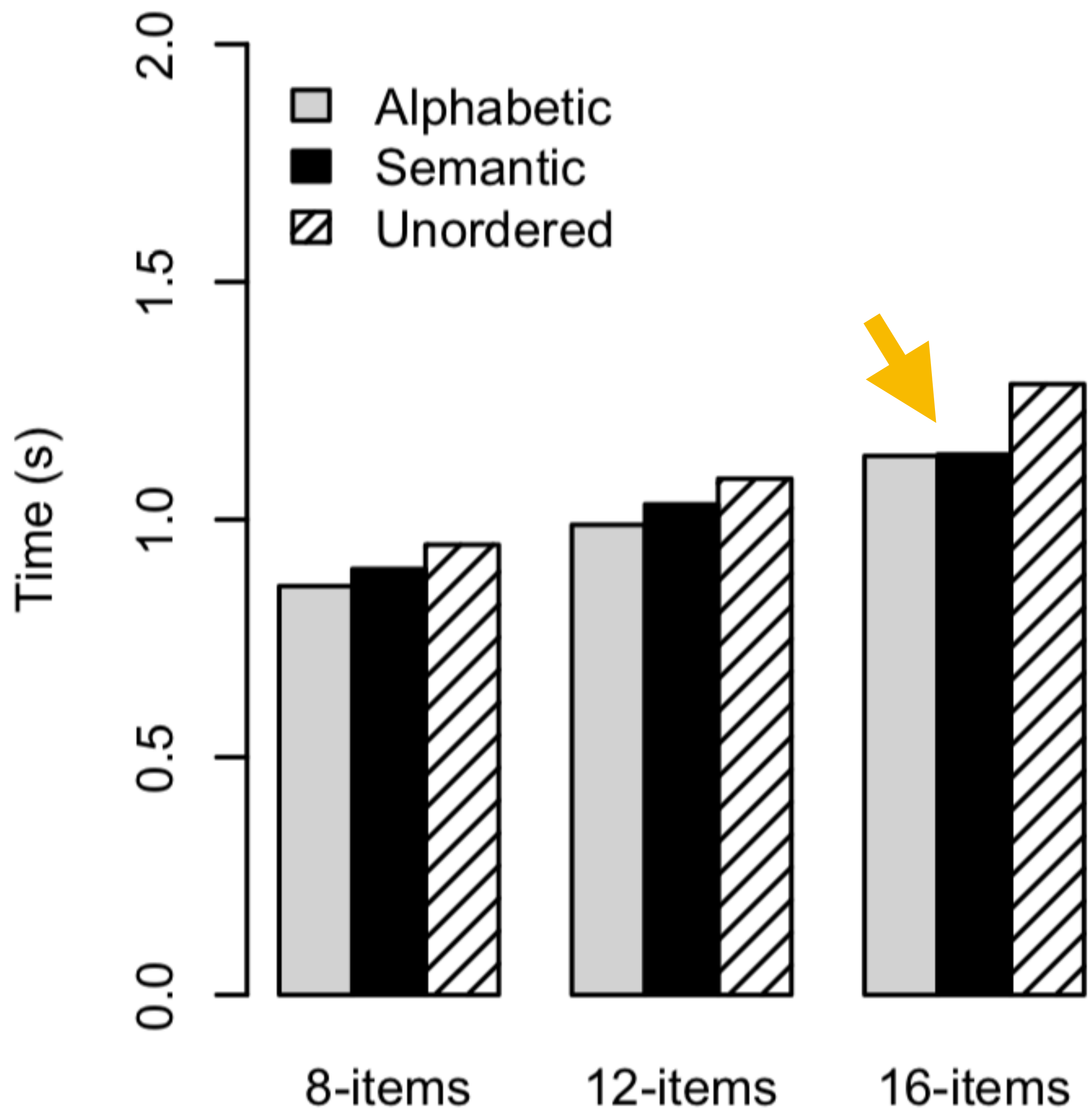
Control error rate under 4 %, such as pointing, text entry, etc.

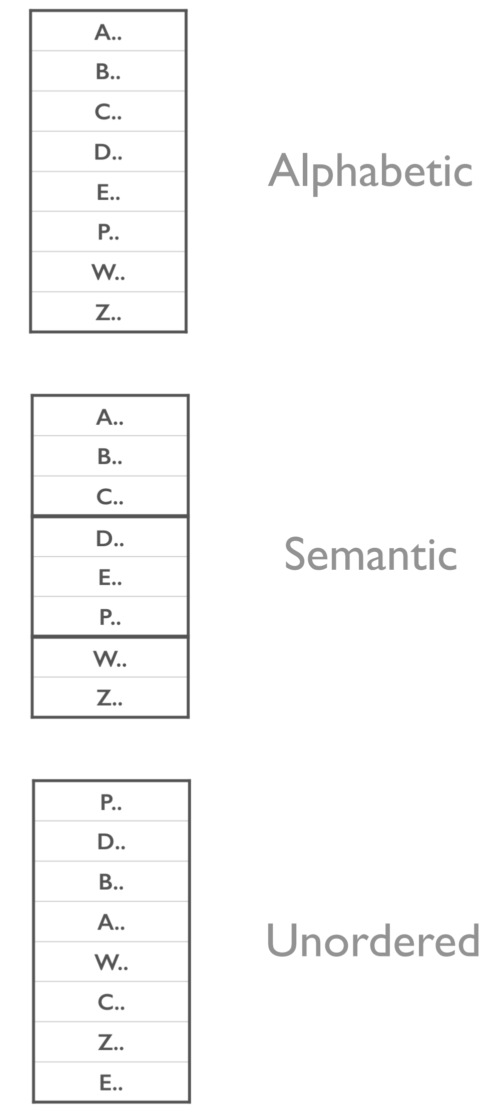
Solution: Control errors

Control error rate under 4 %, such as pointing, text entry, etc.

Remove errors from data analysis.

Measures





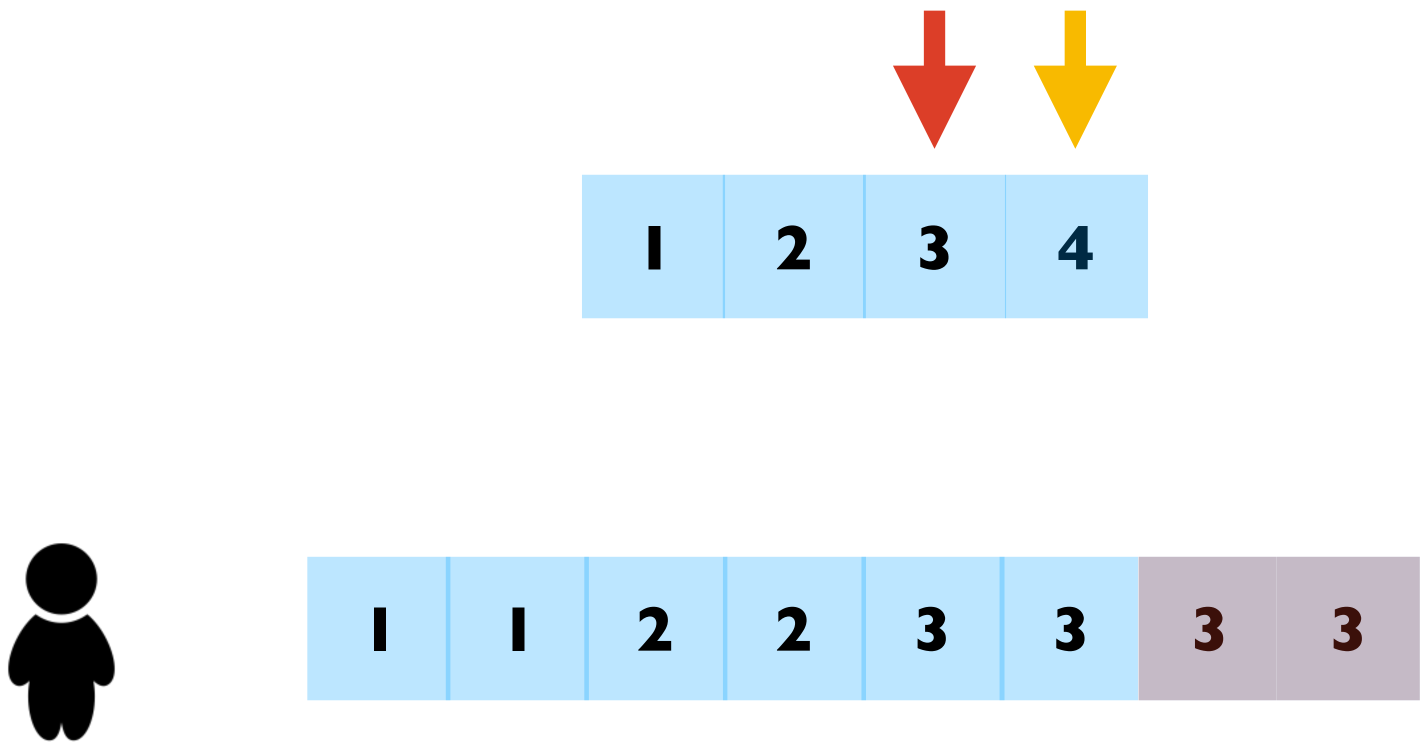
Model of visual search and selection time in linear menus. (Bailly 2014)

Drawback 2: The treatment of errors

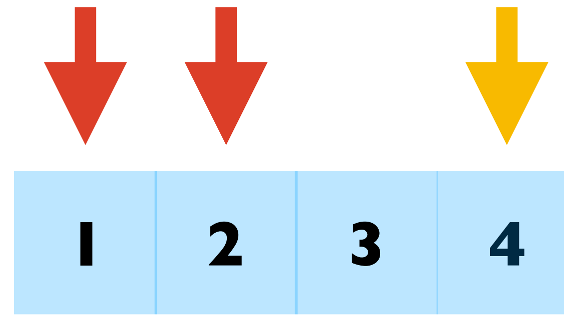
1	2	3	4

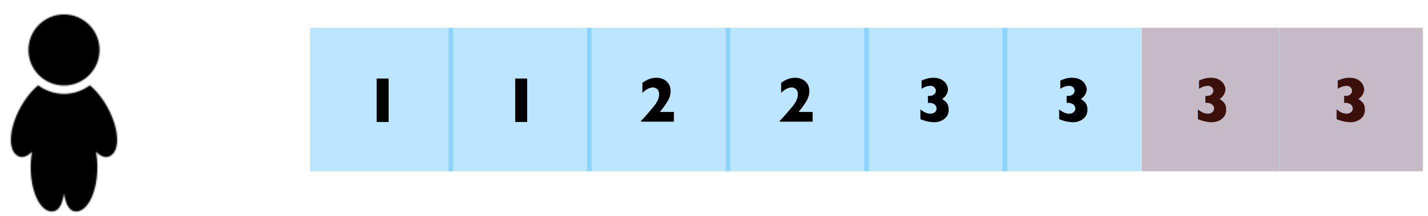
1	2	3	4

1	1	2	2	3	3	4	4

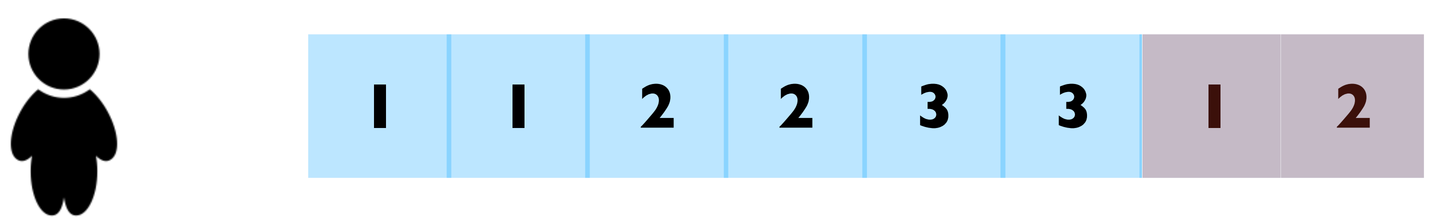


Error rate: 2 / 8 = 25 %





Error rate: 2 / 8 = 25 %



Error rate: 2 / 8 = 25 %

Because human-computer interaction studies a human and a machine in **communication**, it draws from supporting knowledge on both the machine and the human side.

User

communication

Computer

ACM SIGCHI Curriculum for human-computer interaction (2009)

Part iii: Information-theoretic Measures

Measures

User **Computer**

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

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

X takes values in

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

Input entropy: How much information could be transmitted.

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

1	2	3	4

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

Y: The actual input received by the computer.

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

Y: The actual input received by the computer.

1	2	3	4

1	1	2	2	3	3	4	4

$$H(X) = 2 \text{ bits}$$

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

It describes how much information actually gets transmitted.

1	2	3	4

1	1	2	2	3	3	3	3

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

User **Computer**

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

It describes how much information actually gets transmitted.

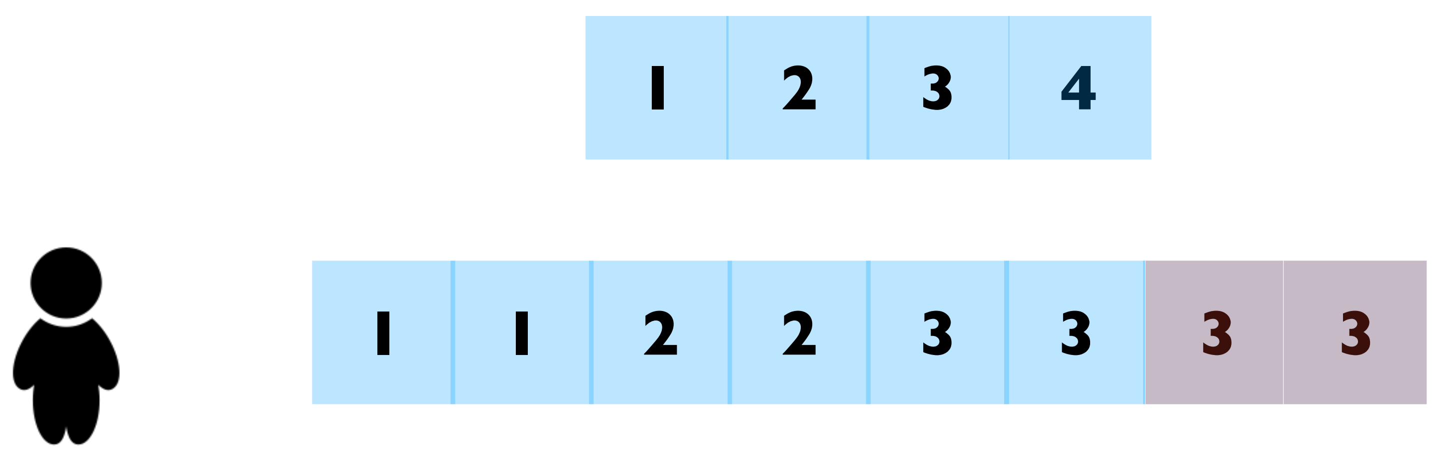
1	2	3	4

1	1	2	2	3	3	3	3

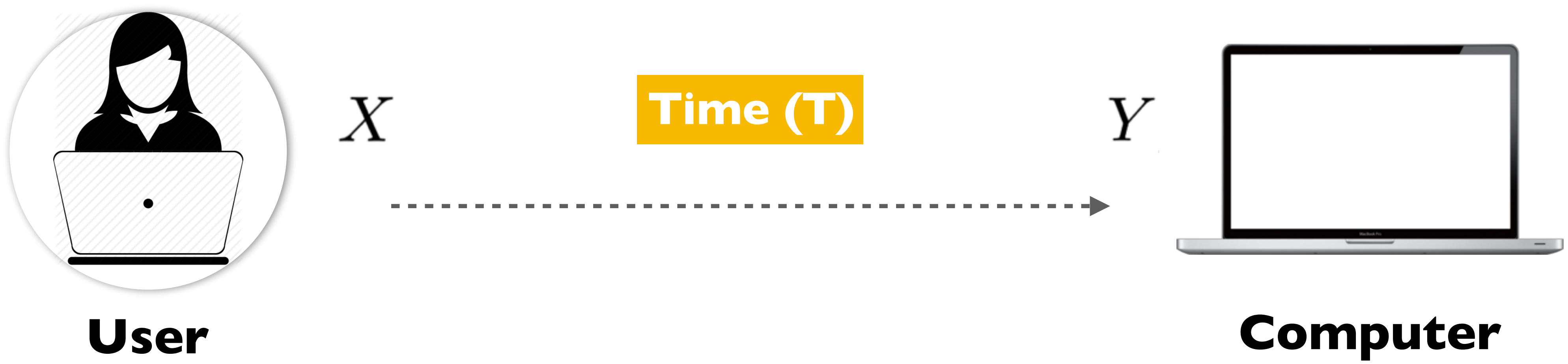
$$I(X;Y) = \sum_x \sum_y P(X=x, Y=y) \log_2 \frac{P(X=x, Y=y)}{P(X=x)P(Y=y)}$$

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

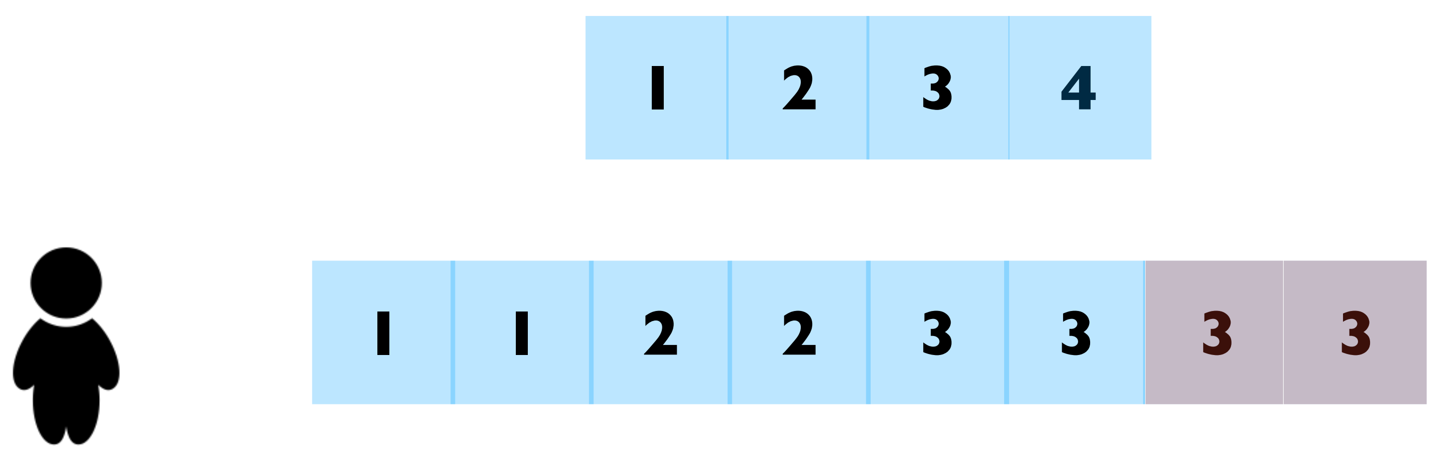
It describes how much information actually gets transmitted.



$I(X;Y) = 1.5$ bits



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



$$TP = I(X;Y)/T = 1.5/1.5 = 1 \text{ bit / s}$$

X

Time (T)

Y

User

Computer

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

1	2	3	4

1	1	2	2	3	3	4	4

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

Advantage 1:

Standard language to describe 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.

Advantage 2:

More consistent measure

Advantage 2:

More consistent measure

	$H(X)$
Time	
Input size	
Distribution	
Error rate	

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$
Time		
Input size		
Distribution		
Error rate		

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$
Time			
Input size			
Distribution			
Error rate			

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP_i
Time				
Input size				
Distribution				
Error rate				

Advantage 2:

More consistent measure

	H(X)	I(X;Y)	H(X|Y)	TP_i	TP_m
Time					
Input size					
Distribution					
Error rate					

$$TP_m = ID/MT$$

$$MT = a + b \times ID$$

Fitts' throughput and the speed-accuracy tradeoff (Mackenzie 2008)

Advantage 2:

More consistent measure

	H(X)	I(X;Y)	H(X|Y)	TP_i	TP_m	TP_z
Time						
Input size						
Distribution						
Error rate						

$$TP_z = 1/b$$

$$MT = a + b \times ID$$

Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT (Card 1978)
Characterizing computer input with Fitts' law parameters - the information and non-information aspects of pointing (Zhai 2004)

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP_i	TP_m	TP_z
Time $\nearrow$						
Input size						
Distribution						
Error rate						

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP_i	TP_m	TP_z
Time $\nearrow$	$-$	$-$	$-$			
Input size						
Distribution						
Error rate						

Advantage 2:

More consistent measure

	$H(X)$	$I(X;Y)$	$H(X\|Y)$	TP_i	TP_m	TP_z
Time $\nearrow$	$-$	$-$	$-$	$\searrow$	$\nearrow$	$-$
Input size						
Distribution						
Error rate						

Advantage 2:

More consistent measure

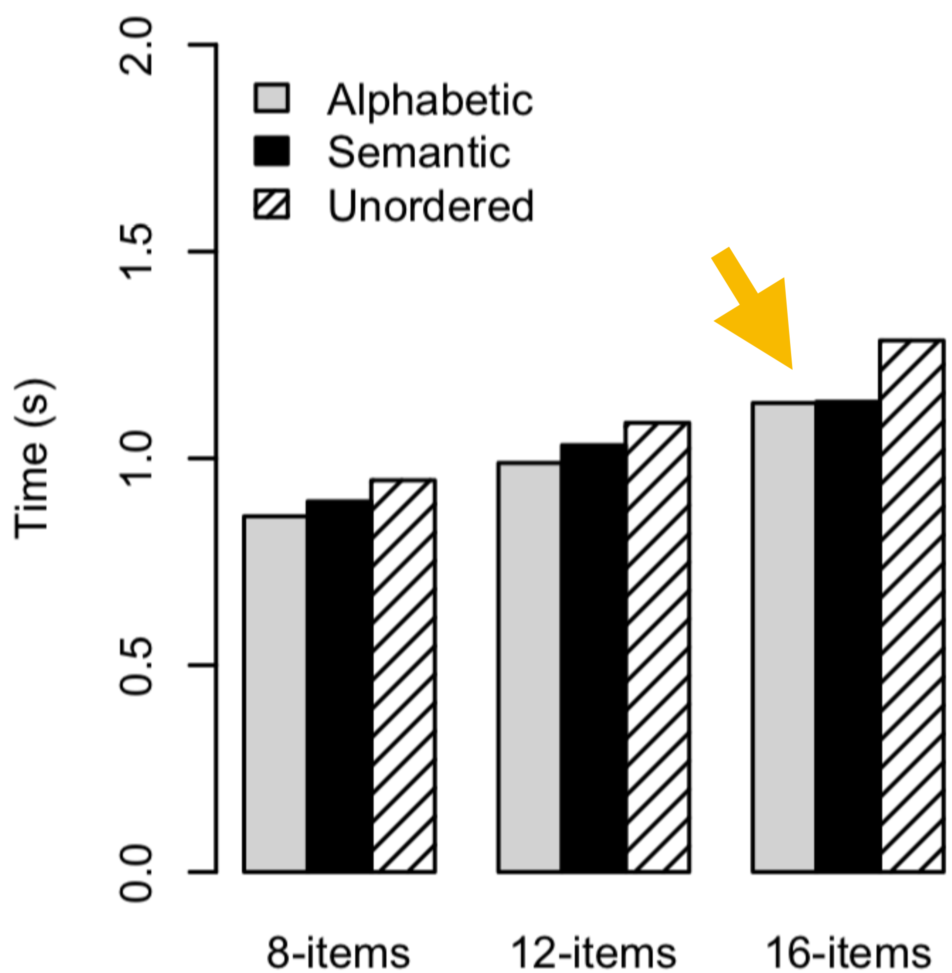
	H(X)	I(X;Y)	H(X|Y)	TP_i	TP_m	TP_z
Time	–	–	–	$\searrow$	$\nearrow$	–
Input size	$\nearrow$	$\nearrow$	–	$\nearrow$	$\nearrow$	–
Distribution	$\searrow$	$\searrow$	–	$\searrow$	–	–
Error rate	–	$\searrow$	$\nearrow$	$\searrow$	$\searrow$	–

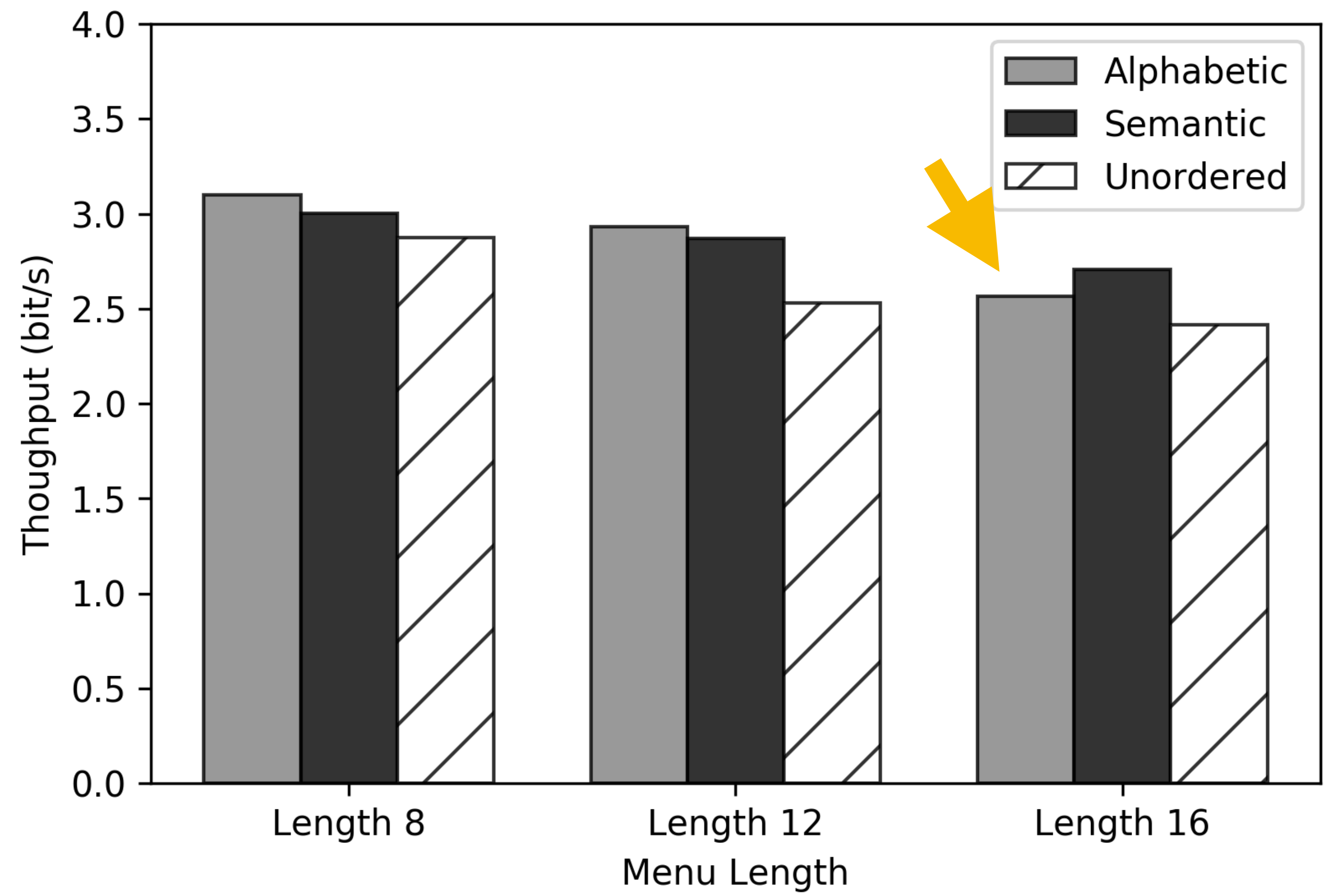
Advantage 3:

Speed-accuracy tradeoff

Advantage 3:

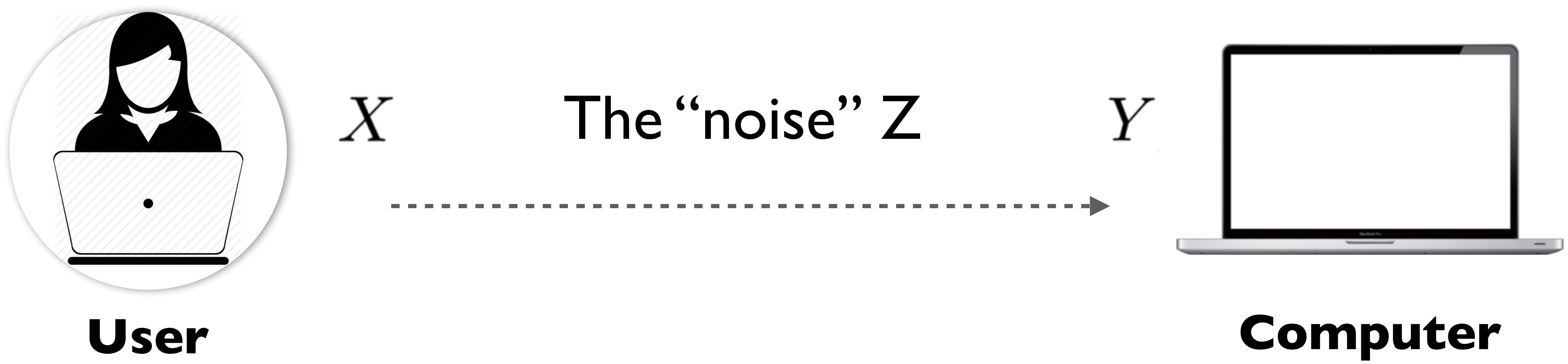
Speed-accuracy tradeoff





Advantage 4:

Equivocation provides information about how users make errors

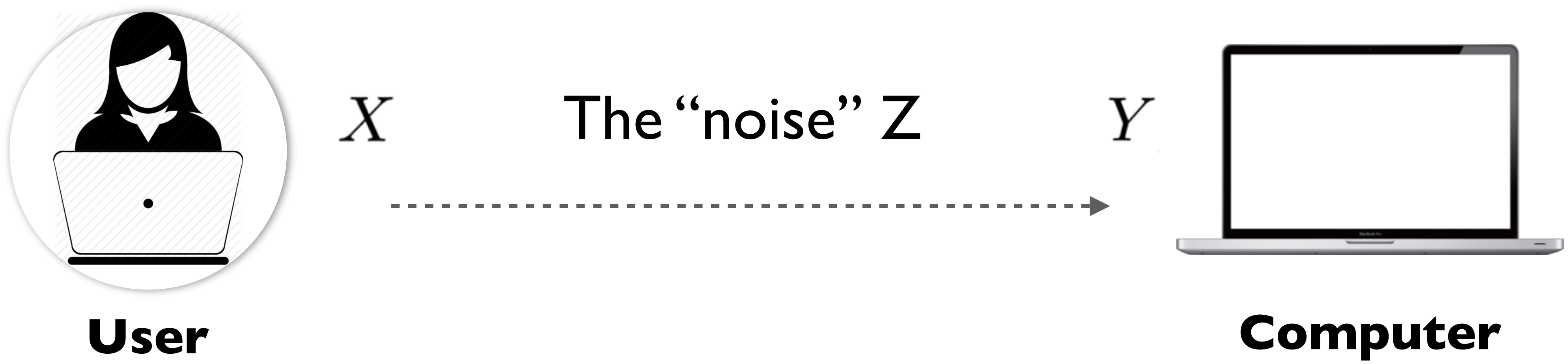


The error random variable:

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

The error rate:

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



The error rate P_e has binary entropy:

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

X The "noise" Z Y

User

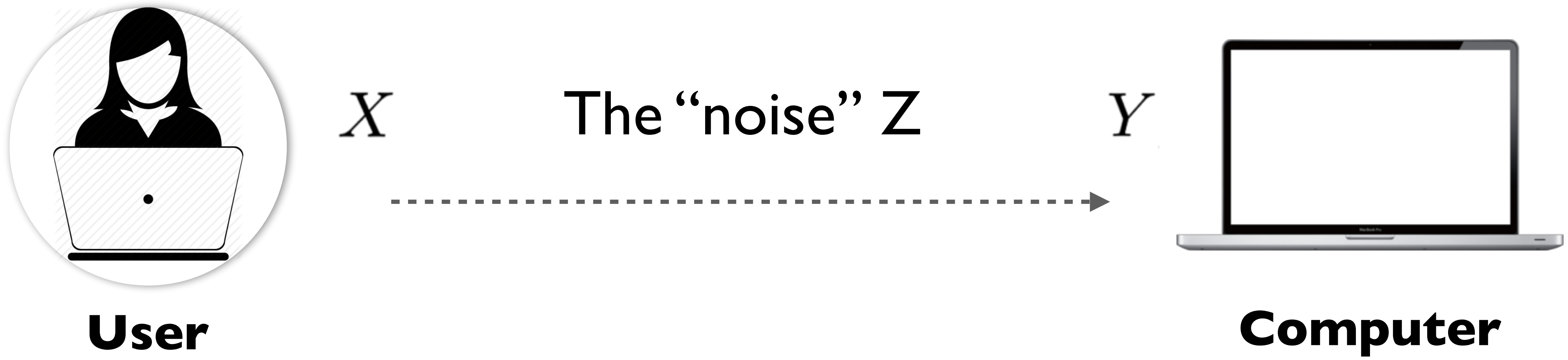
Computer

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

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

Fano's inequality.

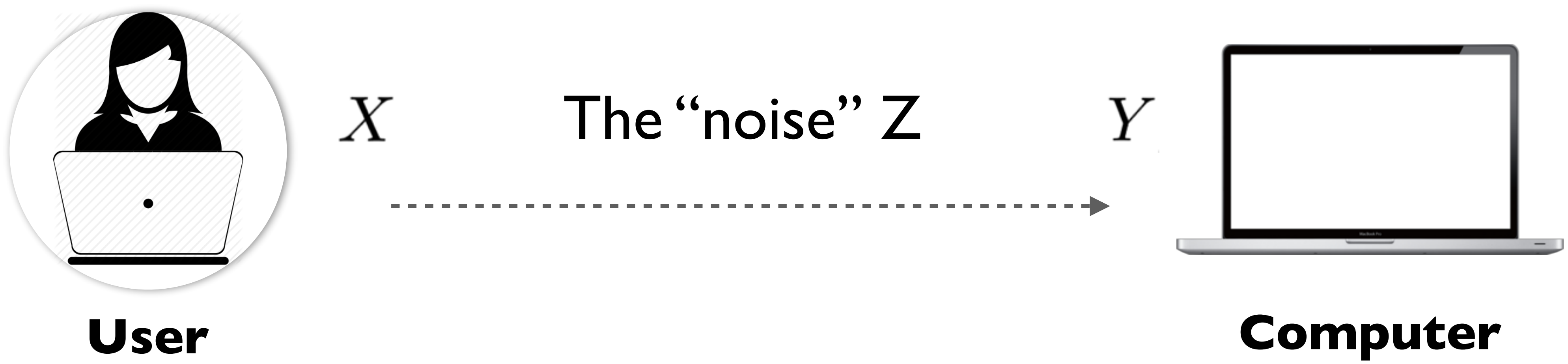
[Theorem 2.10.1] *Elements of information theory*. Cover, T. M., & Thomas, J. A. (2012).



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

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

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



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

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

How they make errors.

Measures

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

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

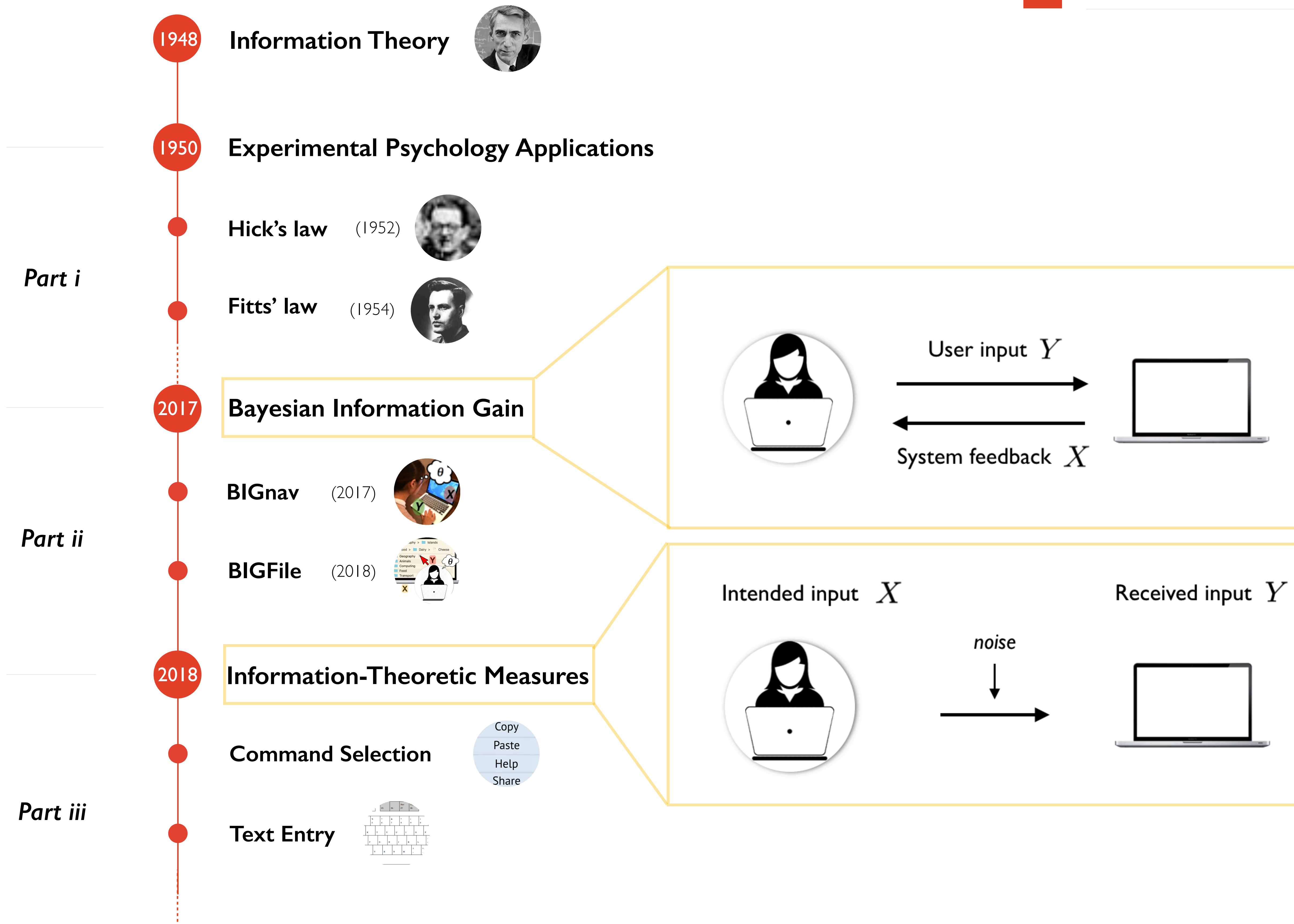
$= 1.3/1.5 = 0.8$ bit / s

Error rate: 2 / 8 = 25 %

Information-theoretic Measures

Wanyu Liu, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard.
Information-Theoretic Analysis of Human Performance for Command Selection. (INTERACT '17).
Wanyu Liu, Antti Oulasvirta, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard.
Information-theoretic Measures for Characterizing Interaction (TOCHI '19) [under preparation]

Contribution



Other “BIG” applications

Bayesian Information Gain for Guiding Design Exploration
(in collaboration with Luana Micallef, University of Copenhagen)

BIGnote: Bayesian Information Gain for Collaborative Music Composition
(came up with the idea during lucid dreaming) August 10, 2018, 12:46 AM

- **Other "BIG" applications**

- **Refine "BIG" framework**

  Model real user behavior $P(Y = y | \Theta = \theta, X = x)$
  User independent
  Machine learning
  Shared control
  Maximize other utility function

- Other “BIG” applications
- Refine “BIG” framework
- Information-theoretic measures for other interaction tasks

  Continuous input
  Taking advantage of equivocation

Main publications

Wanyu Liu, Rafael Lucas D'Oliveira, Michel Beaudouin-Lafon, and Olivier Rioul.
BIGnav: Bayesian Information Gain for Guiding Multiscale Navigation. (CHI '17).

Wanyu Liu, Olivier Rioul, Michel Beaudouin-Lafon, and Yves Guiard.
Information-Theoretic Analysis of Human Performance for Command Selection. (INTERACT '17).

Wanyu Liu, Olivier Rioul, Joanna McGrenere, Wendy Mackay, and Michel Beaudouin-Lafon.
BIGFile: Bayesian Information Gain for Fast File Retrieval. (CHI '18).

Merci

#HumanComputerInteraction #MutualInformation #Bayesian #ComputationalInteraction #CoAdaptation

Thank you everyone!

Merci à tous!

Concert party

- **Prediction algorithm: AccessRank**

$$w_n = {w_{m_n}}^{\alpha} {w_{crf_n}}^{\frac{1}{\alpha}} w_{t_n}$$

AccessRank: predicting what users will do next (Fitchett & Cockburn 2012)

Prediction algorithm: AccessRank

$$w_n = \boxed{w_{m_n}}^{\alpha} w_{crf_n}{}^{\frac{1}{\alpha}} w_{t_n}$$

A Markov chain model

AccessRank: predicting what users will do next (Fitchett & Cockburn 2012)

Prediction algorithm: AccessRank

$$w_n = w_{m_n}{}^{\alpha} \, w_{crf_n}{}^{\frac{1}{\alpha}} \, w_{t_n}$$

A combined recency and frequency model

AccessRank: predicting what users will do next (Fitchett & Cockburn 2012)

✻ Prediction algorithm: AccessRank

$$w_n = {w_{m_n}}^{\alpha} {w_{crf_n}}^{\frac{1}{\alpha}} w_{t_n}$$

A time weighting model

AccessRank: predicting what users will do next (Fitchett & Cockburn 2012)