Quantitative epistemology

Readings for today

- Dretske, F. I. (1983). Précis of Knowledge and the Flow of Information. Behavioral and Brain Sciences, 6(1), 55-63.
- Vlastelica M. (2019). Learning Theory: Empirical Risk Minimization. Towards Data Science.

Topics

1. What is data science?
2. Information flow & knowledge
3. Data science as epistemology
4. Class overview

What is data science?

The story of data

$$x_i$$

The story of data

$$y_i \leftarrow x_i$$

The story of data

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \leftarrow \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} \\ x_{2,1} & \cdots & x_{2,p} \\ \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,p} \end{pmatrix}$$

The story of data

$$Y \leftarrow X$$

The story of data

$$Y = f(X)$$

The story of data

Truth

Experience

Concept class: A set of true functions f that describe the structure of X (and its relationship to Y)

$$Y = h(X) \rightarrow f(X)$$

Hypothesis class: A set of candidate functions h that describe the structure of X (and its relationship to Y)

What is data science?

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.

https://en.wikipedia.org/wiki/Data_science

→

What can I know from my data?

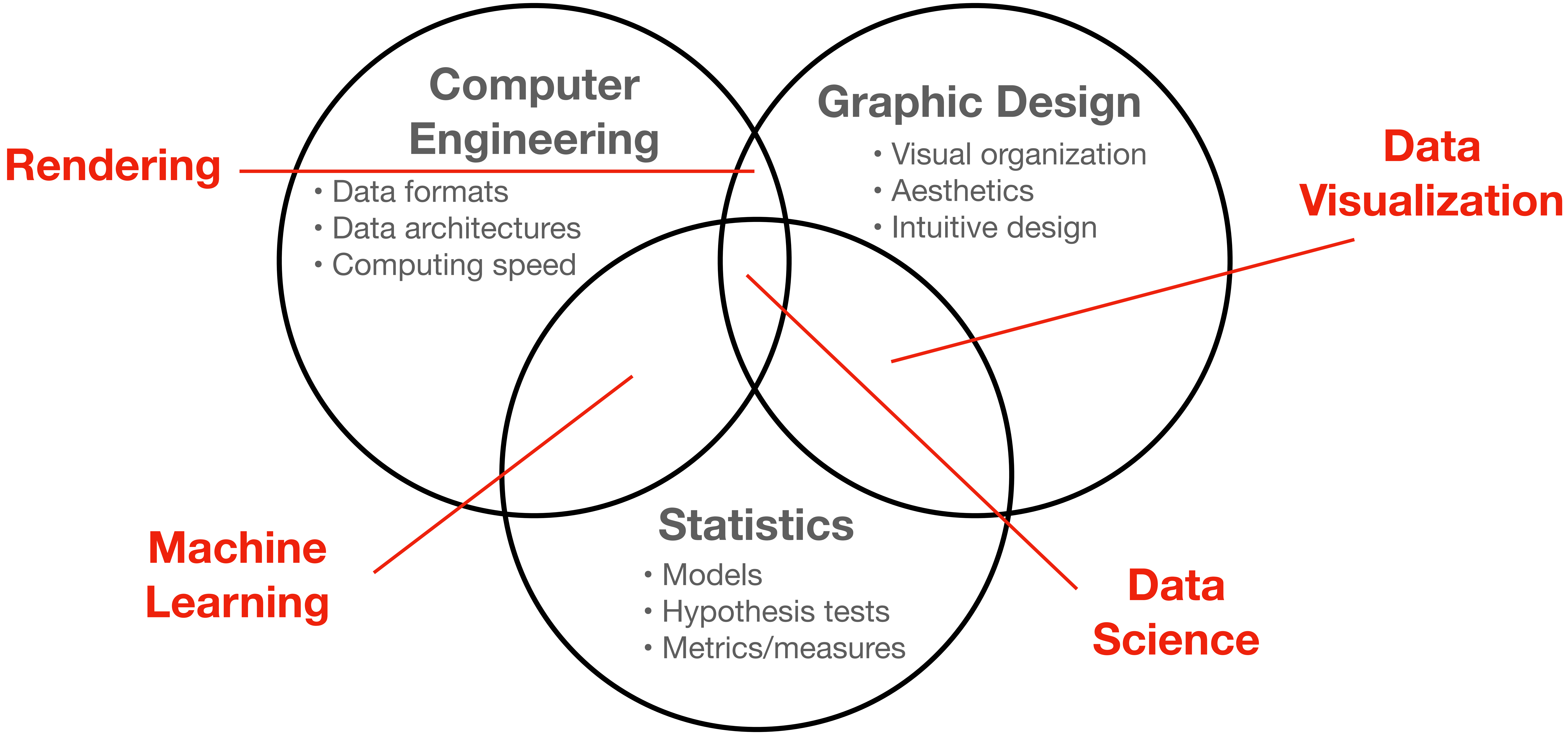
What is data science?

Engineering:
To build

Art & Design:
To communicate

Science:
To understand

What is data science?



Information flow & knowledge

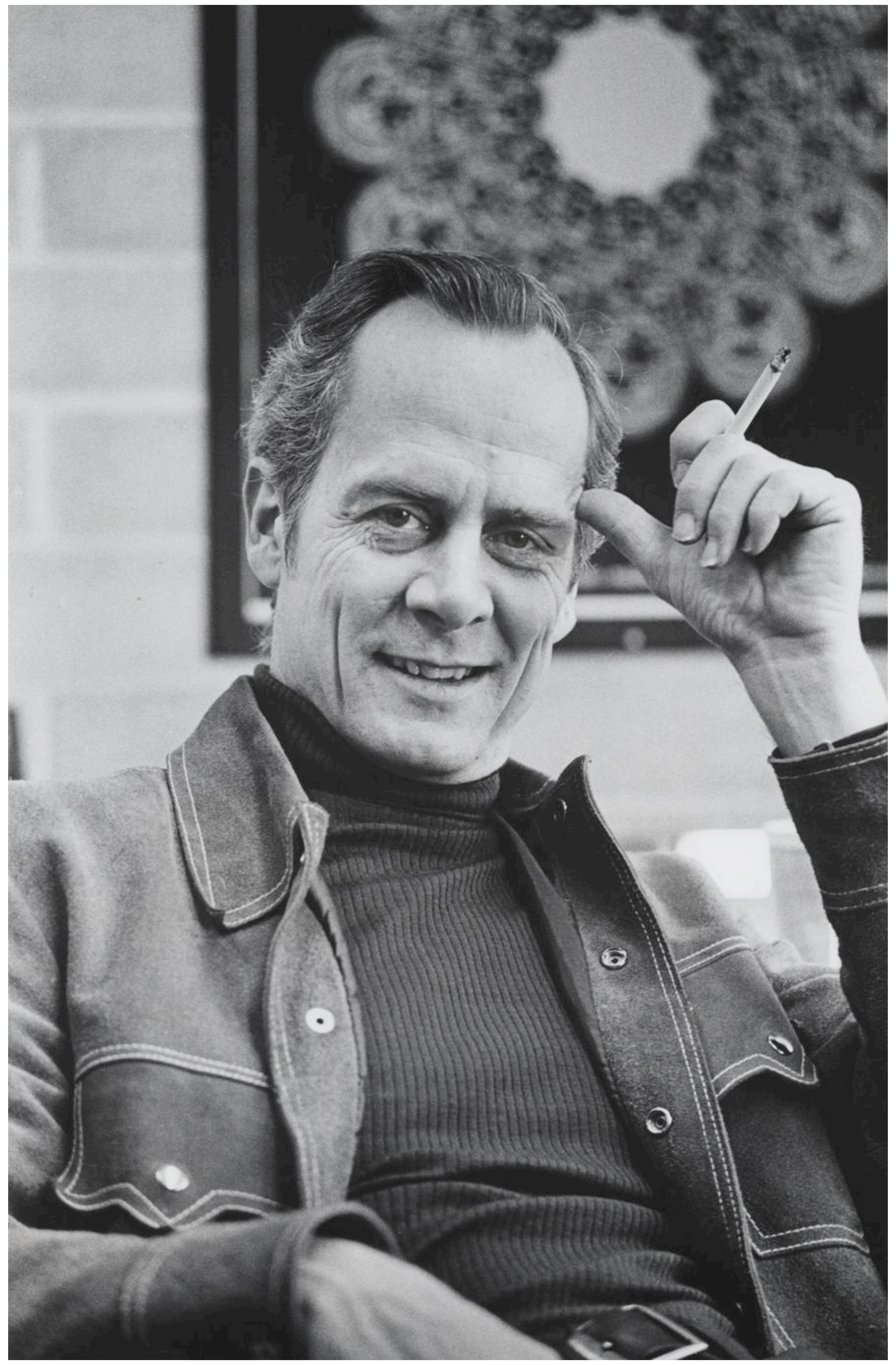
Naturalistic epistemology

Epistemology What is knowledge?

Naturalistic Epistemology Knowledge as a natural kind that is accessible to science

Reliabilism Knowledge arises from the establishment of beliefs based on the reliability of information sources.

Fred Dretske

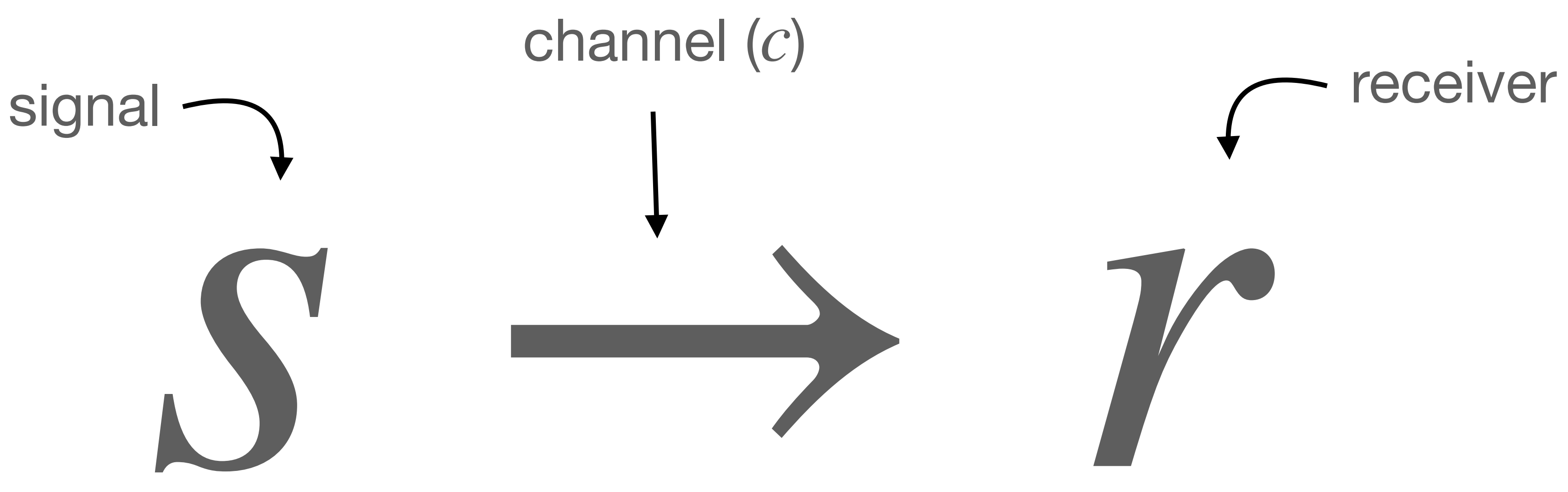


https://search.library.wisc.edu/digital/AI4SINAP6FKSDQ8A

Information theory

Goal: A formal theory for the transmission, processing, extraction, and utilization of information.

Approach: Quantify the *amount* of information a channel, c, can convey about a signal, s, to a receiver, r.



(Shannon & Weaver, 1949)

Amount of information in s

Question: What is the *average* amount of information conveyed by s?

$$I(s) = -\sum p(s_i)\log_2 p(s_i)$$

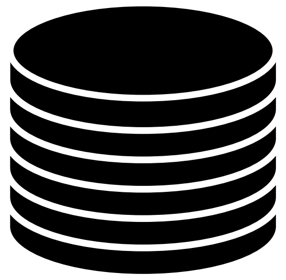
information for signal s

probability that i^{th} state is observed

information available from i^{th} state

$$I(s_i) = log_2(\frac{1}{p(s_i) = 0}) \approx 1$$

Example:



$$I(s) = -\frac{1}{2}log_2(\frac{1}{2}) - \frac{1}{2}log(\frac{1}{2})$$

$$= -\frac{1}{2}(-1) - \frac{1}{2}(-1)$$

$= 1$ bit

(Shannon & Weaver, 1949)

Amount of information received by r

<u>Question</u>: What is the *<u>average</u>* amount of information received by r?

$$I(r) = -\sum p(r_i)\log_2 p(r_i)$$

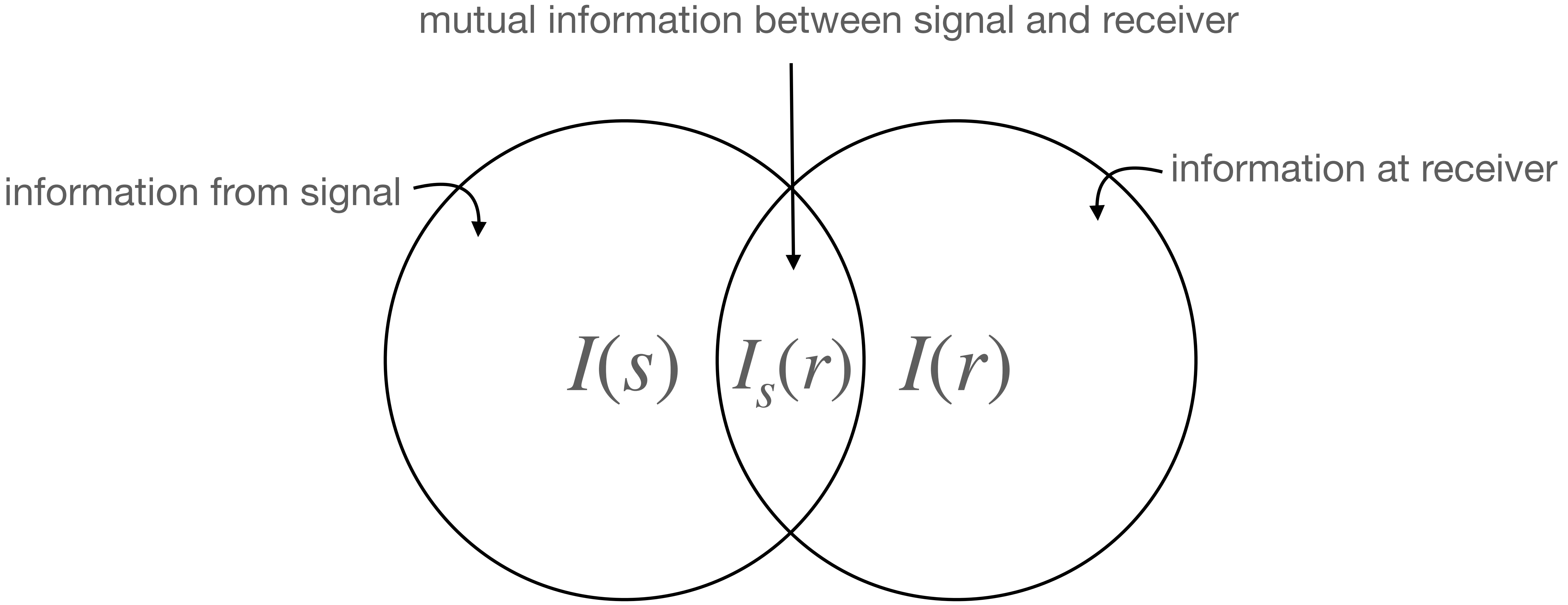
information for receiver r

probability that i^{th} state is observed

information available from i^{th} state

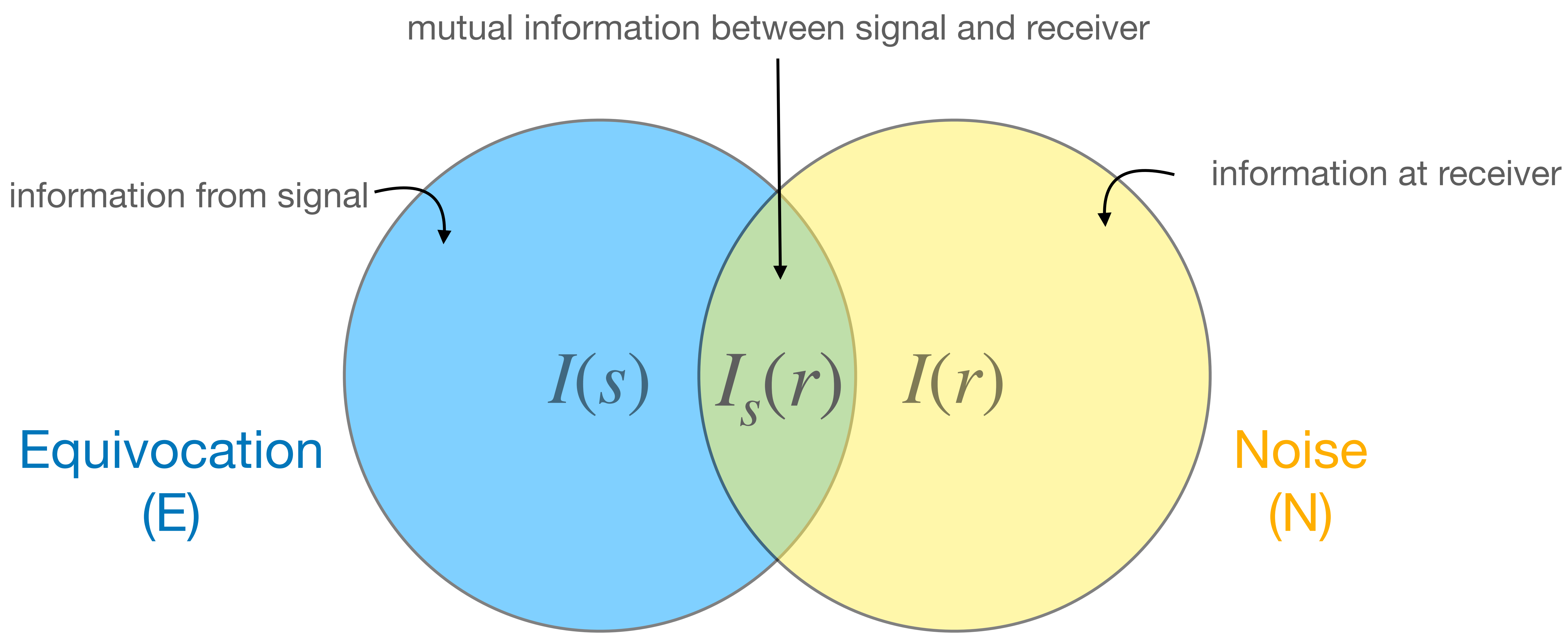
<u>Note</u>: Transmission can change information

$$I(s) \neq I(r)$$



(Shannon & Weaver, 1949)

$I_s(r)$: mutual information



$I_s(r)$: The information transmitted from s to r is the total amount of information available at r, $I(r)$, minus noise.

$$I_s(r) = I(r) - \text{noise}$$

$$I_s(r) = I(s) - \text{equivocation}$$

(Dretske 1981, 1983)

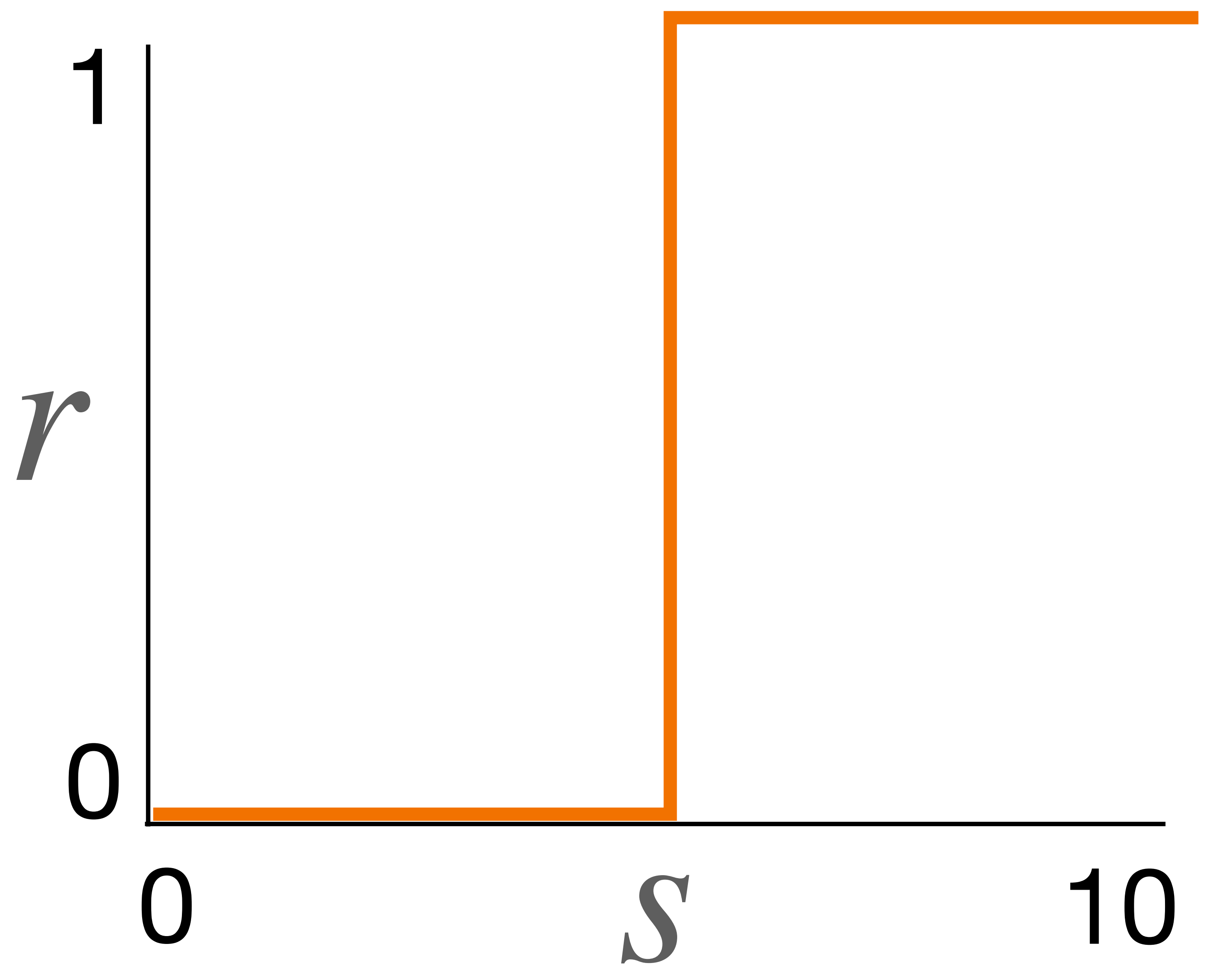
Information theory of meaning

<u>Question</u>: When does signal, s, indicate a specific state of the world, F?

<u>Example</u>:

"$s \geq 5$"

$$\begin{pmatrix} 1 \\ 8 \\ 4 \\ 10 \\ \vdots \\ s_n \end{pmatrix} \rightarrow \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ \vdots \\ r_n \end{pmatrix}$$

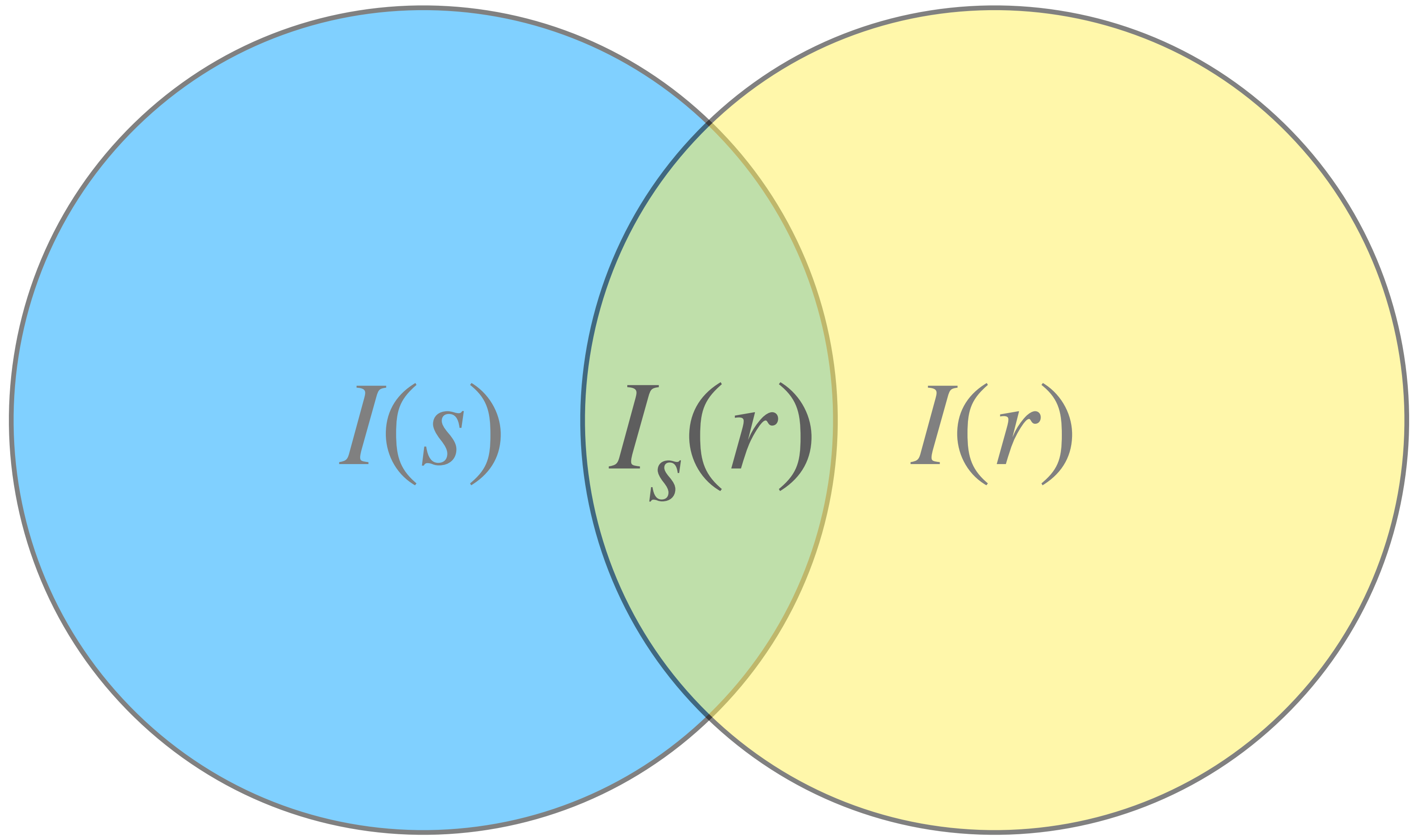


<u>Answer</u>: A signal r (at the receiver) carries the information that s is F if the conditional probability of s's being F, given r (and k), is 1 (but, given k alone, is less than 1)

prior belief

(Dretske 1981, 1983)

Knowledge and the flow of information



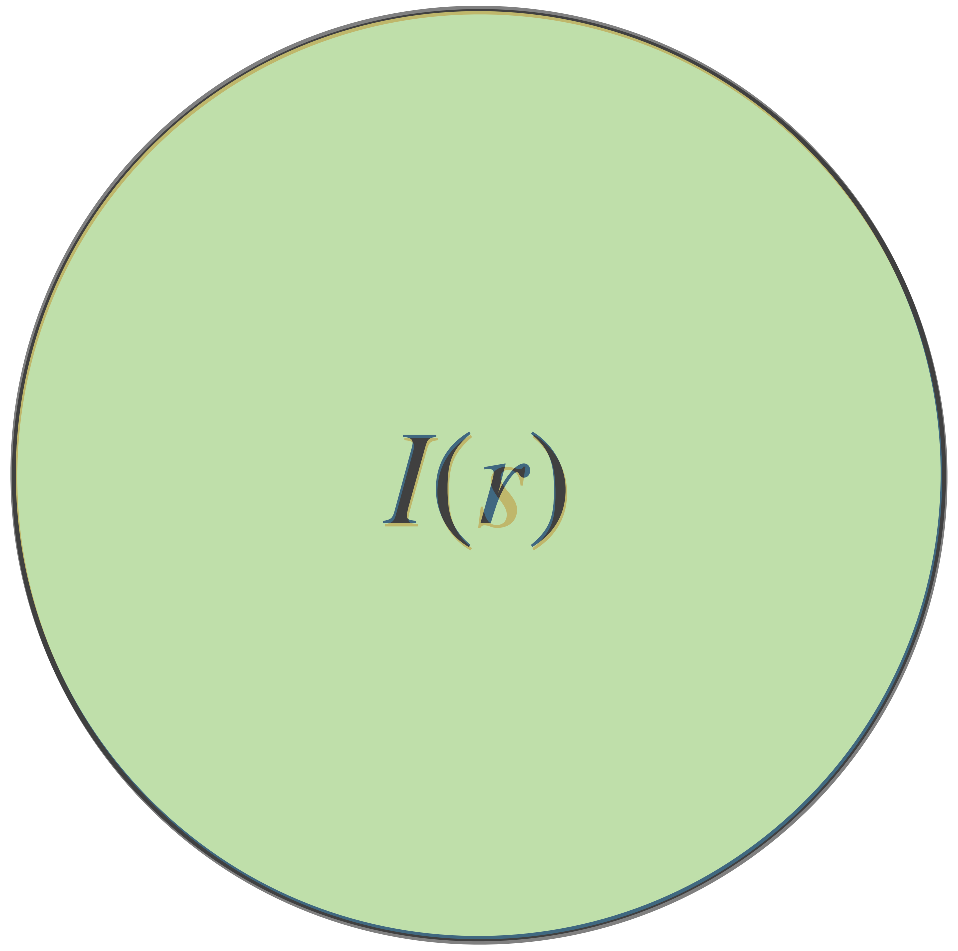
$$s \xrightarrow{I_s(r)} b \longrightarrow a$$

action

(Dretske 1981, 1983)

Knowledge and the flow of information

$I_s(r)$



$I_s(r) = I(s)$ – equivocation $= I(s) - 0 = I(s)$

action

$$s \xrightarrow{I_s(r)} k \longrightarrow a$$

$$h(X) = Y \longrightarrow d$$

hypothesis

data

decision

(Dretske 1981, 1983)

Data science as epistemology

Risk

$$R(h) = \underbrace{\overset{\hat{y}}{\uparrow}\ \ell(h(X), Y)}_{\text{loss function}}$$

Continuous

$$\Sigma(\hat{y} - y)^2$$

Categorical

$$I(\hat{y} = y)$$

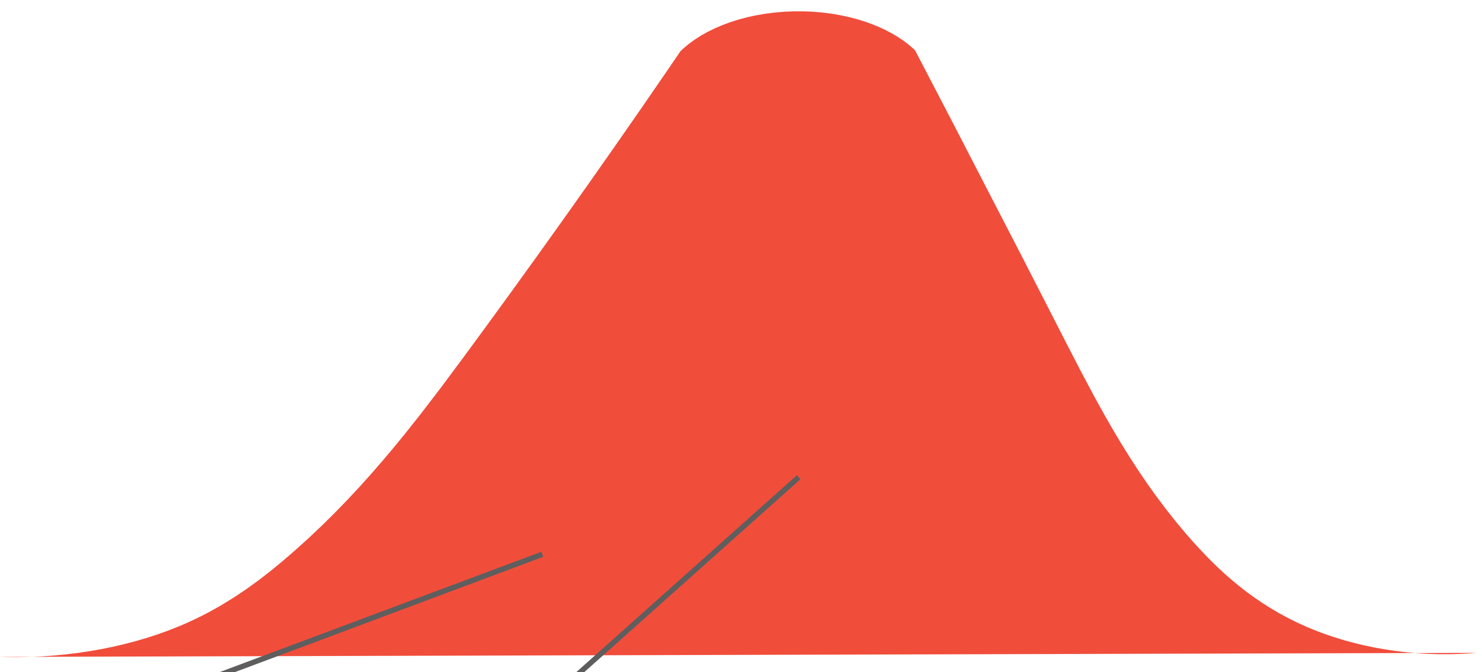
(Vlastelica 2019; Vapnik 1991)

Empirical risk minimization

<u>Expected Risk</u>

$$E_{\text{risk}}\left(h, n, P\right) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{(X,Y)}}_{\text{distribution}}$$

$$= \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{test}} \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{X,Y}}_{\text{distribution}} \underbrace{dP_{(X,Y)}}_{\text{distribution}}$$



<u>Assumption</u>: Both the training and test data come from the same distribution.

(Vlastelica 2019; Vapnik 1991)

Probably Approximately Correct (PAC) Learning

Q: What is learnable?

PAC learning requires a learner to:

1. Approximate the true h
2. Be computationally feasible (P vs. NP)

Empirical risk

Expected risk

$$\underbrace{\mathbb{P}_D[\,|R(\hat{h}) - R(h)| < \epsilon]}_{\text{approximately}} \;\; \underbrace{\geq 1 - \delta}_{\text{probably}}$$

$P_{X,Y}$

(Valiant 1984)

Rethinking what is knowable

Dretske:

A signal r (at the receiver) carries the information that s is F if the conditional probability of s's being F, given r (and k), is 1 (but, given k alone, is less than 1)

Statistical Learning:

A signal Y carries the information that X is F if $Y = f(X)$ is *learnable*.

(Dretske 1981, 1983)

Information → Knowledge → Understanding

Information

$$E_{\text{risk}}(h,n,P) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{X,Y}}_{\text{distribution}}$$

↓

Knowledge

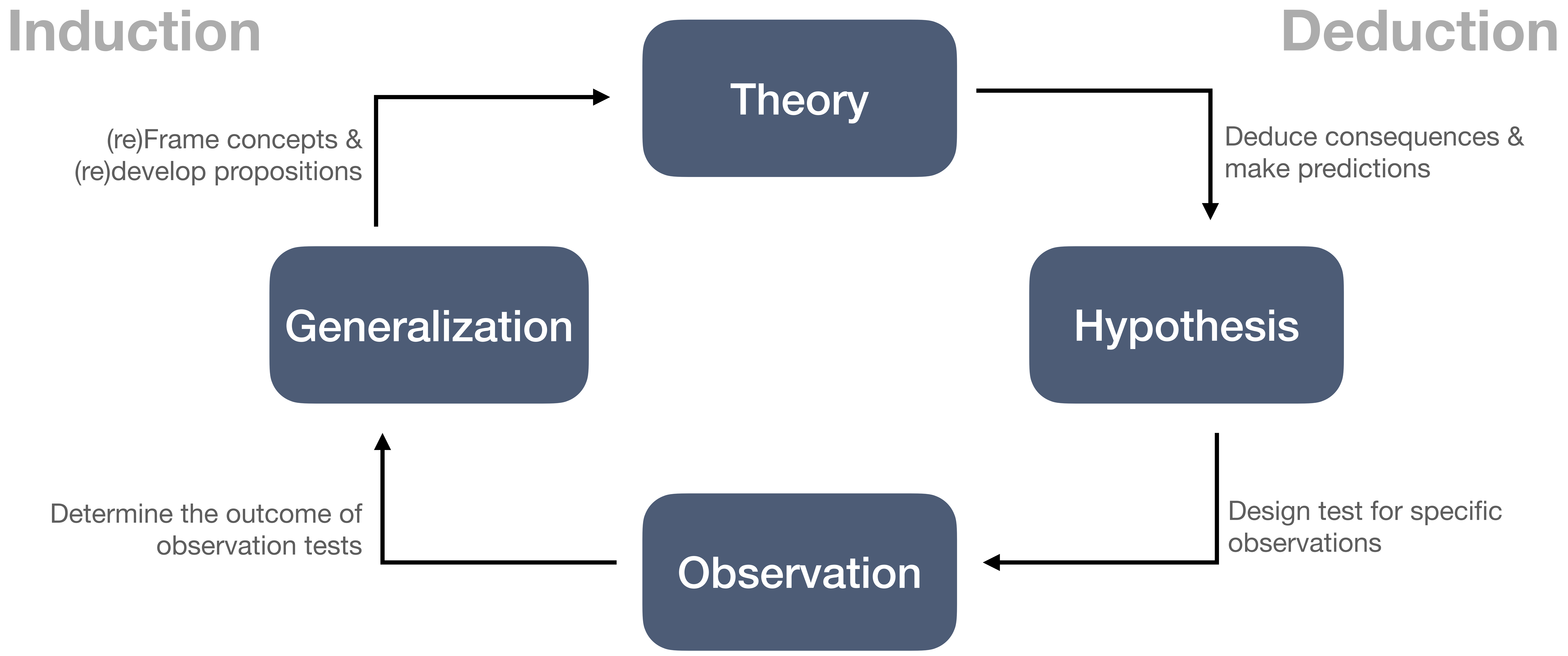
$$E_{\text{risk}}(h,n,P) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{test}} \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{X,Y}}_{\text{distribution}} \underbrace{dP_{(X,Y)}}_{\text{distribution}}$$

↓

Understanding

$$E_{\text{risk}}(h,n,P) = \underbrace{\int_{(\mathbf{X},\mathbf{Y})_n}}_{\text{new}} \underbrace{\int_{(\mathbf{X},\mathbf{Y})}}_{\text{train}} \underbrace{R(h)}_{\text{risk}} \underbrace{dP_{X,Y}}_{\text{distribution}} \underbrace{dP_{(X,Y)_n}}_{\text{distribution}}$$

Hypothetico-deductive model of science



Wallace, W. L. (1971). *The Logic of Science in Sociology*.

(Hempel 1966; Popper 1935, 1963)

Information → Knowledge → Understanding

Information How do we learn the structure embedded in our data?

Knowledge How does the structure in our data predict observations?

Understanding How does our knowledge generalize to new contexts?

Class overview

Goal of the class

Show how data science approaches can be useful for revealing information and knowledge from observational data.

Learning objectives for this class

1. Understand basic principles of statistical theory, measurement, and experimental design;
2. Be able to clean and organize data effectively;
3. Be well versed the execution and interpretation of data analysis;
4. Use information resources to find appropriate data science tools;
5. Communicate statistical results effectively in multiple modalities;
6. Be a critical consumer of data science techniques and their application in empirical research.

Prior knowledge

1. Introductory level understanding of probability theory and statistics (CMU 36-309, 86-309, or equivalent)
2. Basic familiarity with R or similar functional data analysis languages.

Class structure

Reading

Lecture (30-40min)

Offline (asynchronous)

Short Q&A

Breakout Discussions

Open Discussion

Online (synchronous)

Tutorial

Exercises

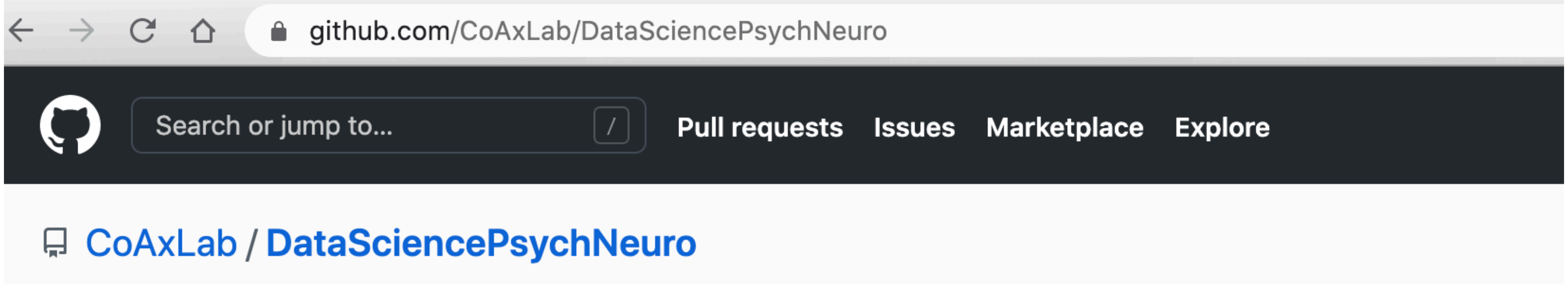
Offline (asynchronous)

Goal:

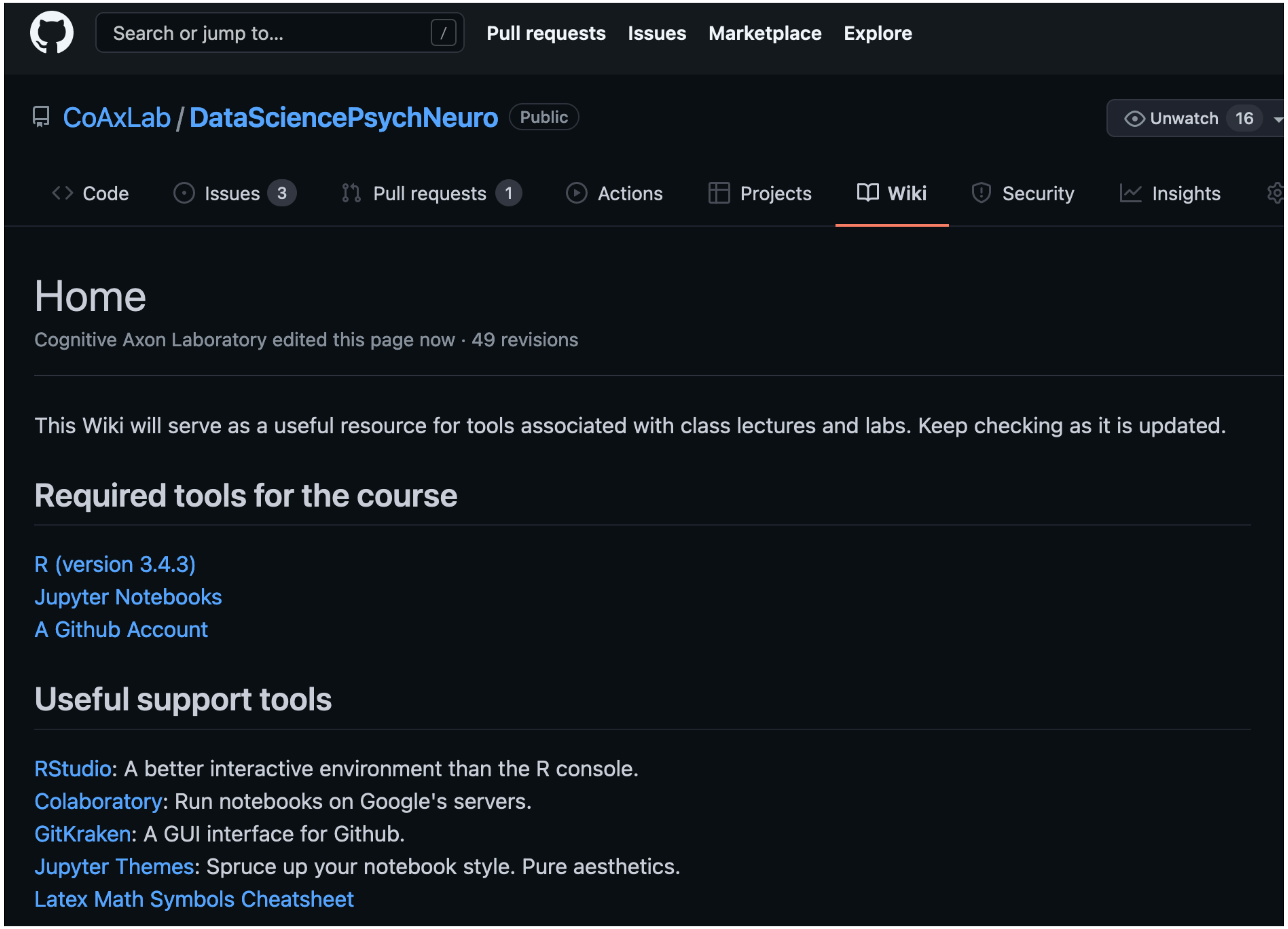
- Content knowledge (crystalized) prior to class.
- Dynamic discussion (fluid) during class.

Resources

1. Texts:
    - Jupyter Book: Data Explorations (https://coaxlab.github.io/Data-Explorations/intro.html)
    - Textbook: James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R (2nd edition). New York: springer. (http://www.statlearning.com).
    - Supplemental book: Hadley Wickham & Garrett Grolemund (2016). R for Data Science. O'Reilly (https://r4ds.had.co.nz/).
    - Auxiliary readings will be posted on Canvas/Github for class sections covering material not in the main textbook.
2. Github Repository: https://github.com/CoAxLab/DataSciencePsychNeuro



Resources


Search or jump to...
Pull requests
Issues
Marketplace
Explore
CoAxLab / DataSciencePsychNeuro
Public
Unwatch 16
Code
Issues 3
Pull requests 1
Actions
Projects
Wiki
Security
Insights
Home
Cognitive Axon Laboratory edited this page now · 49 revisions
This Wiki will serve as a useful resource for tools associated with class lectures and labs. Keep checking as it is updated.
Required tools for the course
R (version 3.4.3)
Jupyter Notebooks
A Github Account
Useful support tools
RStudio: A better interactive environment than the R console.
Colaboratory: Run notebooks on Google's servers.
GitKraken: A GUI interface for Github.
Jupyter Themes: Spruce up your notebook style. Pure aesthetics.
Latex Math Symbols Cheatsheet

Take home message

- Data science can be seen as a branch of epistemology revealing how meaning and knowledge can be determined from information.
- These approaches fit into a larger process of scientific discovery that links abstract theories to empirical data.