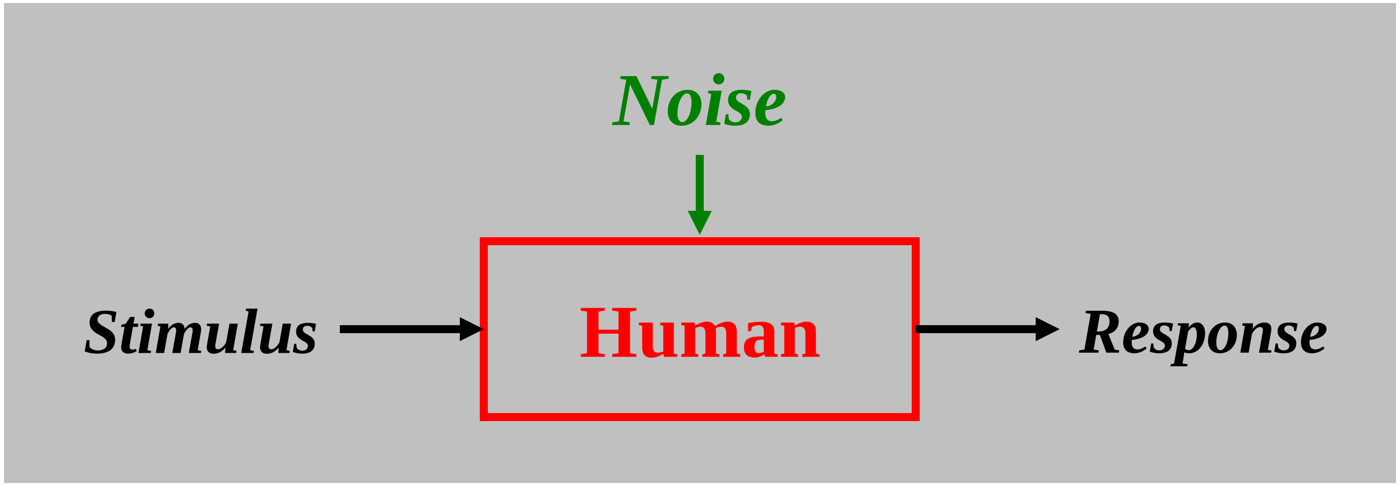
Introduction to Information Theory

Motivation

- Peripheral to central limitations
- Miller's "magic number 7 ± 2"
- Humans as noisy communication channels



Information & Uncertainty

- **Information**
  - **Definition, Condition, Amount**
- **Uncertainty**
  - **An example: number game**
    - **1** **(0 question)**
    - **1 2** **(1 question)**
    - **1 2 3 4** **(2 questions)**
    - **1 2 3 4 5 6 7 8** **(3 questions)**
  - **An *intuitive* measure of uncertainty:**

    $U=\log_2 k$ (k: # of *equally likely* outcomes)

Uncertainty (cont.)

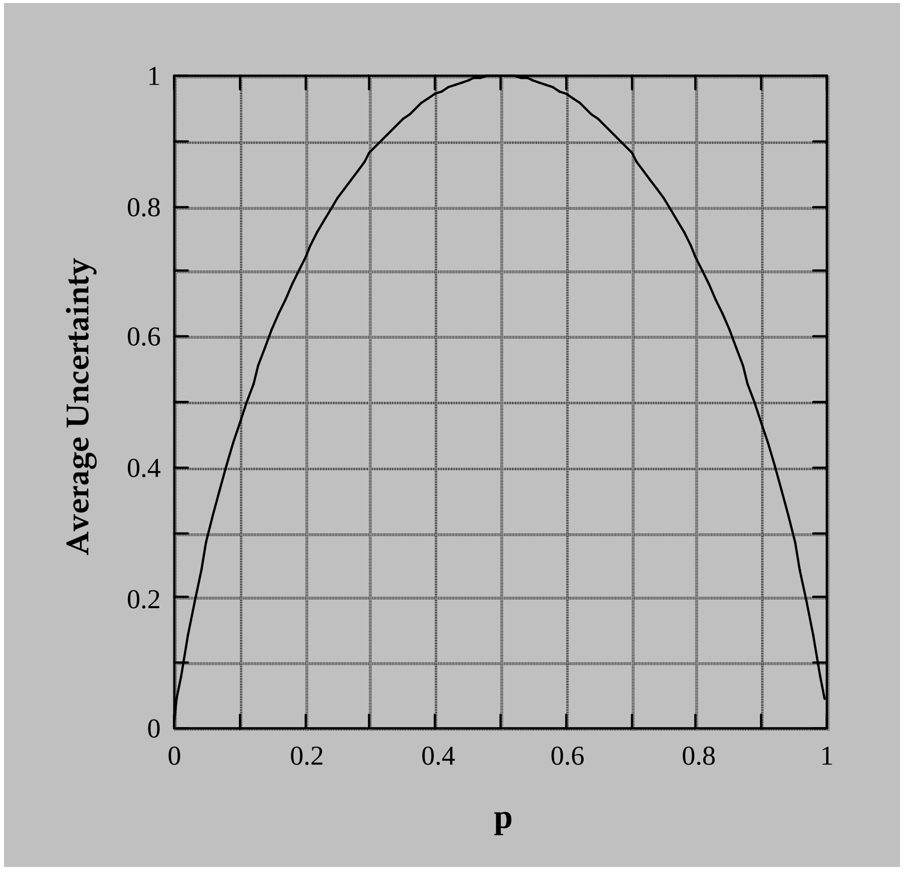
- $U = \log_2 k$ (k: # of *equally likely* outcomes)
- Uncertainty of a given outcome X_i
  - $U_i = \log_2[1/P(X_i)] = -\log_2 P(X_i)$
- If $P(X_i) = 1/k$ (X_i: an outcome, $i \in [1, k]$)
  - Then $U_i = \log_2 k$
- Average uncertainty:
  - $U = \Sigma\, P(X_i)\, U_i$
  - $U = -\Sigma\, P(X_i) \log_2[P(X_i)]$

Average Uncertainty

- Shannon's measure: $U = -\sum P(X_i)\log_2[P(X_i)]$
- Unit for uncertainty and information: *bits*

dichotomous distribution with *p* and *q*=1–*p*

average uncertainty:
$-p\log_2 p - q\log_2 q$



The Absolute Identification (AI) Experiment

Procedures

- **One-interval experiment**
- **Stimuli: S_i, $i \in [1, k]$ ($k > 2$)**
- **Responses: R_j, $j \in [1, k]$**
- **One-to-one mapping ($S_i \Leftrightarrow R_i$)**
- **On each trial, one of the stimuli S_i is presented with an *a priori* probability of $P(S_i)$**
- **Subject makes a response with R_j**
- **Trial-by-trial correct-answer feedback is optional**

S-R Confusion Matrix (e.g., k = 5)

	R_1	R_2	R_3	R_4	R_5	
S_1	14	3	2	0	1	20
S_2	0	13	2	3	1	19
S_3	4	3	11	1	0	19
S_4	2	0	2	15	1	20
S_5	5	3	2	0	12	22
	25	22	19	19	15	100

# of times S_2 was Presented.

# of times the joint event (S_3, R_4) occurred.

# of times R_5 was called.

IS and IR

- **IS (*Information in Stimulus*)**
  - **IS is the average uncertainty in stimulus**

$$IS = -\sum_{i=1}^{k} P(S_i)\log_2 P(S_i)$$

  - **If all stimuli are equally likely, then**

$$IS = \log_2 k$$

- **IR (*Information in Response*)**
  - **IR is the average uncertainty in response**

$$IR = -\sum_{i=1}^{k} P(R_j)\log_2 P(R_j)$$

IT (Information Transfer)

- Also called "mutual information"
- IT = reduction in uncertainty
- For a particular (S_i, R_j) pair:
  - U(S_i) *before*: $-\log_2 P(S_i)$
    - Assuming that P(S_i) is constant throughout the exp.
  - U(S_i) *after*: $-\log_2 P(S_i \mid R_j)$
  - IT(S_i, R_j) = $-\log_2 P(S_i) - [-\log_2 P(S_i \mid R_j)]$

$$IT(S_i, R_j) = \log_2 \frac{P(S_i \mid R_j)}{P(S_i)}$$

- Average IT = $\Sigma\Sigma$ P(S_i, R_j) IT(S_i, R_j)
- IT: the degree of correlation between S's and R's

In-Class Demo

- **Go to "Online Experiments"**
- **Go down to "Part IV. Information Theory"**
- **Select "1. Grayscale Identification"**
- **Click on "Start Experiment"**
- **Select the option with "feedback"**
- **Record the results in your notebook!**

Discussion

- **Compare an absolute-identification experiment with a 1-I 2AFC experiment**
- IS
- IR
- IT
- 2^{IT}

Things to Remember

Uncertainty

- **Uncertainty of a given outcome X_i:**
  $U_i = -\log_2 P(X_i)$
- **Average uncertainty:**
  $U = E[U_i] = -\Sigma P(X_i) \log_2 P(X_i)$
- **Uncertainty for *k* equally likely outcomes:**
  $U = U_i = \log_2 k$

(Memorize!)

IS, IR, IT

- **IS and IR:**

$$IS = -\sum_{i=1}^{k} P(S_i)\log_2 P(S_i)$$

$$IR = -\sum_{j=1}^{k} P(R_j)\log_2 P(R_j)$$

- **IT = reduction in uncertainty**

$$IT = \sum_{j=1}^{k}\sum_{i=1}^{k} P(S_i, R_j)\log_2 \frac{P(S_i \mid R_j)}{P(S_i)}$$

(Memorize!)

Readings

- **Chap. 5: Macmillan, N.A. & Creelman, C.D. (2005). *Detection Theory: A User's Guide*.**
- **Chap. 10: Macmillan, N.A. & Creelman, C.D. (2005). *Detection Theory: A User's Guide*.**
- **G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *The Psychological Review*, vol. 63, pp. 81-97, 1956.**