Information Theory and Psychophysics

- Recall the model of human-machine interaction from the first day. To paraphrase the questions...
  - What information is needed?
  - What information is available? How is it obtained?
  - What does the person do with that information?
  - How do they know what actions to take?
  - What is the result?

Information Theory (pp. 22-24)

- Grew out of the study of problems of electrical communications (especially telegraphy) and statistical mechanics.

- The message source selects one of a possible set of messages, encodes it, and transmits the resulting signal through a channel. The message is decoded and then received by the receiver. Noise can distort or mask the message.

Information Theory: definitions

- **Information**: “reduction of uncertainty” (that the receiver has about the message transmitted by the source.)
- **Bit**: a single unit of information, equivalent to a choice between two alternatives (yes/no, on/off)
- **Entropy (H)**: measure of information in bits (typically, bits per second, bits per word, bits per symbol)
  - "a measure of our ignorance"
  - H is based on the number of possible messages (or events or stimuli) N, the probabilities of those messages, and the context (i.e., the probabilities of all messages).
SO WHAT? Why should we care?

Because if we can quantify information, we can quantify human information processing!! In terms of performance, we can look at choice reaction time (how long it takes to make a choice or decision) as a function of the amount of information.

Example:

A tetris-like game is being designed in which the probability of appearance of each shape is directly proportional to the number of edges. If the pieces are shaped as follows, what is the average amount of information and the % redundancy in the system?

Quantifying information

For N equally likely alternatives,

\[
H = \log_2 N = \frac{\log_{10} N}{\log_{10} 2}
\]

If known probabilities:

\[
H = \log_2 \left( \frac{1}{p_i} \right) = - \log_2 p_i
\]

The *average entropy* of a system:

\[
H_{av} = \sum p_i (\log_2 (\frac{1}{p_i}))
\]

Choice reaction time:

The HICK-HYMAN LAW says that choice reaction time is a function of the amount of information in a stimulus:

\[
RT = a + b H
\]

where,

- \( H \) is the amount of information
- \( a \) is simple reaction time
- \( b \) is the slope of the function (the amount of added processing time for each bit of information.)

Intuitively - the more choices we have, the longer it will take to choose from among them.
Choice reaction time example

In the tetris-like game just described, the simple reaction time and the choice reaction time when all pieces are equally likely is measured. The results are as follows:

- simple reaction time, $a = 300$ msec.
- choice reaction time for equally likely events, $R_{eq} = 800$ msec.

Then the probability of appearance of the individual pieces is changed as described. What is the new expected choice reaction time?

How do we know what humans perceive?

- Psychophysics: How people respond to the strength of physical stimuli of all sorts.
- Greatest concern surrounds the weak and complex stimuli
- Major theoretical landmarks:
  - Classical
  - Signal detection theory
  - Multidimensional scaling
  - Nonlinear psychophysics

Psychophysics – Classical

- Absolute Threshold – minimum value of a stimulus that can be detected
- Difference threshold – minimum change in the value of a stimulus that can be detected.
Two fundamental laws

- **Weber’s Law**
  - \( \frac{\Delta I}{I} = C \)
  - Perception of minimum change is relative to the initial value.
    - \( \Delta I = \) JND
    - \( I = \) initial intensity
  - Ratio is constant for all examples of a type of stimulus.

- **Fechner’s Law**
  - \( \Psi = k \log I \), where
    - \( \Psi = \) psychological intensity
    - \( I = \) physical intensity
    - \( k = \) constant
  - \( k \) is constant for all values of a particular type of stimulus

Classical theories couldn’t account for

- individual differences in sensory acuity,
- motivation factors in response,
- the effect of noise,
- errors in classical theory when confronting extreme stimulus values.
- SDT forms the basis of basic decision theory.
Signal Detection Theory (SDT)

- Absolute threshold — minimum value of a stimulus that can be detected by 50% of the population.
- Difference threshold — minimum change in a stimulus that can be detected by 50% of the population.
- "Just noticeable difference"

A situation is described in terms of two states of the world:
- A signal is present ("Signal")
- A signal is absent ("Noise")

You have two possible responses:
- The signal is present ("Yes")
- The signal is absent ("No")

<table>
<thead>
<tr>
<th>Signal</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hit</td>
<td>False Alarm</td>
</tr>
<tr>
<td>P(H)</td>
<td>P(FA)</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Miss</td>
<td>Correct Rejection</td>
</tr>
<tr>
<td>P(M)</td>
<td>P(CR)</td>
</tr>
</tbody>
</table>

What does this mean?

- If you decide "Yes" and the true state of the world is "Signal," that's called a Hit.
- If you decide "Yes" and the true state of the world is "Noise," that's called a False Alarm.
- If you decide "No" and the true state of the world is "Signal," that's called a Miss.
- If you decide "No" and the true state of the world is "Noise," that's called a Correct Rejection.
Theoretically ...

- The theory assumes that what you are doing is:
  - First, you collect sensory evidence concerning the presence or absence of the signal.
  - Next, you decide whether this evidence constitutes a signal. This means that you must have some criterion C that you use as a "cutoff" (related to the signal strength): if the evidence is less than C, you decide "No"; if the evidence exceeds C, you decide "Yes".

![Diagram](image_url)

Measures of performance in SDT: 1. Sensitivity (d')

- A function of the keenness or sensitivity of the human's detection mechanisms and the relative strength of the signal in noise.
- This value may be calculated from the probabilities of a hit and a false alarm.
- Example: Radiologists evaluating a candidate MRI machine are given 1000 sample slides, 500 of which are from patients with a tumor. The results of the readings are as follows:

<table>
<thead>
<tr>
<th>Tumor (&quot;signal&quot;)</th>
<th>No tumor (&quot;noise&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (&quot;yes&quot;)</td>
<td>350</td>
</tr>
<tr>
<td>Negative (&quot;no&quot;)</td>
<td>150</td>
</tr>
<tr>
<td>100</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>

Calculating d'

- \[ P(H) = \quad \quad z_H = \quad \quad \]
- \[ P(M) = \quad \quad z_M = \quad \quad \]
- \[ d' = \quad \quad \]

![Diagram](image_url)
Measures of performance in SDT:
2. Response bias ($\beta$)

- Another way to describe performance is in terms of response bias: you may be prone to say "yes" (which is "risky") or you may be prone to say "no" (which is "conservative").

- Response bias is the ratio of the heights of the two curves at the cutoff point and is measured by the quantity:

$$\beta = \frac{p(X_i|S)}{p(X_i|N)}$$

where $X_i$ = "evidence variable"
$S$ = signal
$N$ = noise

Going back to our example ...

- $P(F_A) = \underline{\text{__________}}$  $\text{Ord}_{F_A} = \underline{\text{__________}}$
- $P(M) = \underline{\text{__________}}$  $\text{Ord}_M = \underline{\text{__________}}$

<table>
<thead>
<tr>
<th>Tumor</th>
<th>No tumor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (&quot;yes&quot;)</td>
<td>150</td>
</tr>
<tr>
<td>Negative (&quot;no&quot;)</td>
<td>150</td>
</tr>
</tbody>
</table>

$$\beta = \underline{\text{__________}}$$

Some important points ...

- Studies of human performance show that humans do change beta in response to changes in probabilities and payoffs -- but not as much as they should! This phenomenon is called sluggish beta.

- Note: the terms "risky" and "conservative" refer only to a person's propensity to say "yes (signal)" or "no (noise)."

- Examples:
  - radiologists reading x-rays for signs of tumors
  - radar operators on a battle ship looking for incoming enemy aircraft
  - scanning a parking lot for a parking space
Important points (cont.)

- The cutoff (C) for determining the presence of a signal is not the same as the response bias parameter (b). However, they are correlated. For example, if your strategy becomes more “risky” (in other words, your b goes down), then your cutoff also goes down. As you get more conservative, both your cutoff and your b go up.

- What is an “optimal setting” for b?
  Note that as terms in the denominator increase, b decreases and thus response becomes riskier. As terms in the numerator increase, b increases and thus responses become more conservative. Also, the first fraction shows the effect of signal likelihood and the second fraction shows the effect of payoffs.

\[ b_{opt} = \frac{P(\text{Noise})}{P(\text{Signal})} \times \frac{\text{Value(CR)} \times \text{Cost(FA)}}{\text{Value(Hit)} + \text{Cost(Miss)}} \]

- The receiver Operating Characteristic (ROC) curve plots the probability of a hit against the probability of a false alarm. Each curve represents the same sensitivity at different levels of response bias.

Stevens Power Law

- \( \Psi = k \ (C_s - S_0)^C \) where
  - \( \Psi \) = psychological rating of signal strength
  - k = constant to account for idiosyncrasies in the experiment, (not connected to the type of signal)
  - \( C_s \) = physical signal strength
  - \( S_0 \) = absolute threshold for the signal (using the 50% rule)
  - C = constant for all examples of a particular type of signal, e.g.
    - Loudness of a sound, 0.6
    - Brightness of a light, 0.33
    - Smell of coffee, 0.55
Performance can also be affected by “vigilance” requirements

- E.g., radar operators. Performance degrades
- After extended hrs of operation
- After extended sleep loss
  - 21 hrs awake = BAC 0.08%
- With simultaneous tasks > serial tasks

Remedies

- Auditory cueing
- Cognitive tasking
  - New task with high cog demands actually has a refreshing effect

Stochastic Resonance

- Sometimes a small amount of noise, when added to a weak signal, increases the chance of detecting the signal.

Fuzzy SDT

- Conventional: correct response = [0,1]
- Fuzzy: Odds of a correct response
  - $s = \text{probability of a stimulus matching the target signal}$
  - $r = \text{priority of a response}$
- Incorporates certainty judgments
  - Hit (H) = min($s, r$); Hit rate = $\Sigma(H)/\Sigma(S)$
  - Miss (M) = max($s-r, 0$); Miss rate = $\Sigma(M)/\Sigma(S)$
  - False alarm (FA) = max($s-r, 0$); FA rate = $\Sigma(FA)/\Sigma(1-S)$
  - Correct rejection (CR) = min($1-s, 1-r$); CR rate = $\Sigma(CR)/\Sigma(1-S)$
- Distinguishes near-miss situations from clear-cut cases.
Multidimensional Scaling

- Stimuli with multiple dimensions
- How similar are the stimuli? (Sights, sounds, tastes, smells?)
- Analysis of matrix of similarity judgments produces the number of dimensions and
- Coordinates on each dimension for each stimulus.
- Classic application: Perception of similarity of chromatic colors will full saturation.