Assigned project presentation order

- Goal of the midterm presentation is to have a dialog with the class about the project. If you're not here, you miss out on that dialog. That's why it's so important to be here.
- How do we want to divide up presentations? Random numbers to order projects?
- Went through the list of projects to see who would and would not be here
  - Groups missing on Thursday: Emotional salience, Driving simulator, brain-computer interface robot, flashing icons, perceived attractiveness (one person here on Thursday)
- The class agreed that Emotional Salience, Driving Simulator, Brain-Computer Interface, and Flashing Icons will go on Tuesday, everyone else will go on Thursday
- During presentations, five to seven minutes for presentation, then some time for questions. Next group needs to setup during the questions so they can start immediately.

Topic for today's class: Psychophysics
The goal of psychophysics is to make objective measurements of subjective experiences. The term psychophysics derives from *psycho* (e.g. mental processes / subjective experiences) and *physics* (measurement).

Physical properties, such as *length*, *weight*, or *intensity*, can be measured directly. Physical properties can generally be paired with a subjective, experiential property, such as *size*, *heaviness*, and *brightness*. While all observers will generally agree on physical properties, they may disagree on subjective properties. Two observers may each agree that an object weighs 25 lbs, but one may consider this a heavy weight while the other considers it a light weight. Accordingly, we cannot measure subjective properties directly, but instead rely on responses given to us by other people.

*Ex. Consider size in virtual reality. The perceived size or distance of objects can vary greatly depending on several factors. Heaviness depends on a lot of haptics features. Graphics depends a lot on brightness.*

Psychophysical methods are a set of tools that help quantify the relationship between physical properties and subjective properties. Psychophysical methods have their origins in psychology, but they are widely applied to other fields that intersect with technology. There are two primary psychophysical quantities/methods: thresholds and scales.
Thresholds
There are two types of thresholds: absolute thresholds and difference thresholds. Absolute thresholds measure the intensity at which a specific stimulus becomes detectible to an observer (e.g. the volume at which a sound becomes audible.) Difference thresholds measure the point at which a difference can be detected between two stimuli. The difference threshold can also be referred to as a “just noticeable difference”.

When dealing with color/intensity differences, it is common to ask people to compare two colored squares, or one colored square against a colored background. This example will be referenced frequently in this section.

Shown below is an example of a task measuring the absolute threshold at which a gray square can be distinguished from a black background. The three images represent three different states of the system where the color of the square is varied.

![Gray square against black background](image1)

Shown below is an example of a task measuring the difference threshold at which the color of two squares can be distinguished from each other. The three images represent three different states of the system where the color of the square is varied. This type of task can also be referred to as a *discrimination task*.

![Colored squares](image2)
Scales

While thresholds measure the point at which two values can be distinguished, scales measure the perception of relative relationships between two values. For instance, at what point is a light perceived as twice as bright as another light? At what point is one weight perceived as twice as heavy as another weight?

History of Psychophysics

Much of the seminal research in psychophysics was done by two researchers in the 1800’s: Weber and Fechner.

Weber wanted to understand heaviness. He wanted to explore if people could tell the difference between two weights. He was one of the first to quantify the experience of lifting a weight. He had them lift one weight then another weight and asked if the weights were different (a discrimination task). The interesting finding was that the just noticeable difference was proportional to the weight itself. The threshold increases as the weights increase. This became known as Weber’s law: \( \Delta \alpha = K \alpha \). The just-noticeable difference of an intensity is proportional to the intensity itself.

Question: Is there a universal \( K \) for all modalities? For all individuals?

Answer: \( K \) is different for different modalities, but is pretty consistent for different individuals.

Fechner was asking a similar question: how do we measure sensations. He connected measurements in physics to measurements of subjective sensations: in order to measure subjective sensations, you need a zero and a unit to increment by, just like you do in the physical world. Examples of this include measuring temperature in celcius or weight in kilos.

For a subjective experiences, the zero is defined as the detection threshold and the incremental difference is defined as the just noticeable difference. This means the incremental difference changes as you change the intensity. Ex: |-||--|--|----|--------|-----------------

Fechner also introduced methods for measuring threshold. One of the most common methods is called the method of adjustment. This is the method that would be followed in the previous examples of how the different thresholds could be detected.

- Select a standard value (e.g. a square of color 128).
- Take a square of a different value (e.g. 135) and ask an observer to adjust this squares color until it matches the standard. This adjustment can be performed using a dial. Turn the dial until the color matches the standard.
  - In this case, you turn the dial down until you reach the desired color. This is called a descending trial
- Record the start intensity and the final intensity (e.g. 135 and 126)
- Repeat another trial, but now begin using a color with lower intensity (e.g. 115).
  - In this case, you dial up. This is called an ascending trial.
- Note the state and final intensity. (e.g. 115 and 125)
- Repeat until the desired amount of data is collected (20+ normally).
- You get data like this

<table>
<thead>
<tr>
<th>Trial #</th>
<th>Series Type</th>
<th>Original Intensity</th>
<th>Final Intensity (ascending)</th>
<th>Final Intensity (descending)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>108</td>
<td>126</td>
<td>---</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td>135</td>
<td>---</td>
<td>124</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>110</td>
<td>127</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>140</td>
<td>---</td>
<td>123</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>100</td>
<td>124</td>
<td>---</td>
</tr>
<tr>
<td>6</td>
<td>D</td>
<td>133</td>
<td>128</td>
<td></td>
</tr>
</tbody>
</table>

- Next compute the mean of final intensity (ascending) and final intensity (descending). If these are different, it means you have a bias in your experiment.
- Finally, compute the grand mean. This is the point of subjective equality (PSE). This is the intensity which is subjectively equivalent to our standard value of 128.

**Question: Why is this useful?**
Imagine you make LCD displays, and your LCDs have some amount of noise within it. If the noise is less than the PSE, then the noise doesn't matter. You can determine the PSE using the above type of testing.
Now imagine you make paints. There’s no point to make paints that lie within their own intervals of uncertainty, as people won’t be able to tell them apart.

The just noticeable difference (JND) can be calculated as follows: 0.67*standard deviation of the PSE distribution.
The Upper level (UL) threshold at which you can see a difference is $PSE + JND$
The Lower level (LL) threshold at which you can see a difference = $PSE - JND$
The interval of uncertainty is $UL - LL$

**Signal Detection Theory**

Signal detection theory relates to when a signal can be detected. Theoretically, a signal moves from undetectable to detectable immediately after crossing the threshold. In practice, this detection occurs probabilistically, and thus the exact point at which the signal is detected varies for different observers. These two viewpoints are reflected in the image below. Practically speaking, you have an S shaped curve, not a sharp boundary.
Consider the example of a radar operator looking for enemy aircraft. He watches the radar screen for blips, which could represent enemy aircraft, or could be due to noise. The bell curves on the graph below represent how the probability of a blip being due to either noise or an actual aircraft, depending on the intensity of the signal. The curve on the left represents noise due to the system and to the observer’s senses. The curve to the right represents a real event.

The observer has to decide whether a particular signal is due to noise or an enemy aircraft. This decision can be made by establishing a decision threshold. Two decision thresholds are shown on the graph, marked by the red and blue line. Each threshold comes with pros and cons. The red threshold will generate very few false positives, because very little of the noise curve extends to the right of the threshold. However, this threshold will miss several true signals, because the event curve extends to the left of the threshold. The opposite is true of the blue threshold, which will have many false positives, but will rarely miss an actual event. How the system should be calibrated depends on the relative cost of the two types of mistakes.
Stimulus response matrix

The above explanation can also be represented using a stimulus response matrix (see below). The contents of each cell define the relative probability of each combination of true or false signal and reaction to it.

<table>
<thead>
<tr>
<th>Stimulus present (e.g. enemy aircraft)</th>
<th>Response present (e.g. sounds alarm)</th>
<th>Response not present (e.g. does nothing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus present</td>
<td>Hits P(Yes</td>
<td>Present)</td>
</tr>
<tr>
<td>Stimulus not present (e.g. no aircraft)</td>
<td>False positive/alarm P(Yes</td>
<td>Not present)</td>
</tr>
</tbody>
</table>

Question: For the stimulus response matrix, are the miss sections and the false positive sections equivalent to Type I and Type II errors?
Answer: Yes, they are equivalent.

The actual probabilities of these events can be calculated over multiple trials.

Signal detection theory allows us to quantify both the true sensitivity and biases in what we are measuring. True sensitivity reports the difference in means between signal and noise; the larger the difference, the more sensitive your equipment/observer. Bias gives us a sense for how likely it we are to make a miss. Both of these equations are reported below.

True sensitivity = $d' = \mu_{su} - \mu_N / \sigma_N$

Response bias = $\beta = \gamma_S/\gamma_N$

$N = \text{noise}, S = \text{signal}, \mu = \text{mean}, \sigma = \text{standard deviation}$

$\gamma = \text{point at which the decision threshold intersects the two probability distributions}$

Two alternative false choice task (2AFC)

More recently, the two alternative false choice task (2AFC) has been developed to reduce bias in threshold recognition tasks. Rather than allowing users to adjust a signal until it appears to match a comparison value, participants are shown two signals and asked which matches the comparison value more closely. For example, people could be shown two colored squares (127 and 131) and are asked which is most similar to the comparison square (128). Participants are required to make a choice, even if they think both squares look the same.

This task format allows you to reduce bias inherent in the “dial” type of task, because it eliminates the possibility of extreme values. A user cannot set the dial to 150 and say it looks equal to the square of value 128.
In this type of task, users are shown a range of different signals. The threshold can be identified once users are selecting between the two options at random. Once this has occurred, the values being shown to the user are inside of the point of subjective equality.

**Additional Example**

Imagine you want to determine the degree of photo manipulation required before a user can detect that the image has been manipulated (e.g. an algorithm that smooths people’s skin). This could be determined using a 2AFC. Show participants pairs of images, one of which is unmanipulated and one of which has been manipulated. The detection threshold can be determined by identifying the point at which users begin selecting the manipulated image at a rate greater than 50% (the rate due to chance).